UNIVERSIDAD DE MURCIA

FACULTAD DE BIOLOGÍA

Genetic variability in mosses and its relation to climate change adaptation processes in Mediterranean environments

Variabilidad genética en musgos y su relación con procesos adaptativos al cambio climático en ambientes Mediterráneos

D. Mahmoud Magdy Abdallah Awad

2013

UNIVERSIDAD DE MURCIA

FACULTAD DE BIOLOGÍA

Genetic variability in mosses and its relation to climate change adaptation processes in Mediterranean environments

Thesis to obtain the Ph.D. title, presented by

Mahmoud Magdy Abdallah Awad

M.Sc. Murcia University, B.Sc. Ain Shams University

This study was realized in the Department of Plant Biology of Murcia University, under the direction and supervision of Dr. Olaf Franziskus Werner and Dr. Rosa María Ros Espín

Dr. Olaf Franziskus Werner
Ph.D. Heidelberg University

Dr. Rosa María Ros Espín
Ph.D. Murcia University

Murcia, March 2013

To my family,
especially Samah, my wife

ACKNOWLEDGMENTS

The present study has been done with the intensive help and guidance of Dr. Olaf Franziskus Werner and Dr. Rosa María Ros Espín. Due to their great discussions, exchange of intellectual information, comments, critics, kindness and friendship, I will always be in debt for their great help and support.

I am also grateful for the personal and scientific comportment of my lab mate and best friend Sergio Pisa Martín and his family to my life during the last four years in Murcia.

I owe an especial gratitude to Dr. Bernard Goffinet (Connecticut University, U. S. A.) for sharing valuable information, which was of great impact to the current work. Many thanks to Dr. Adnan Erdağ (Menderes University, Turkey), Dr. Helena Hespanhol (CIBIO, Research Center on Biodiversity and Genetic Resources, Porto, Portugal), Mr. Heribert Köckinger (Weisskirchen, Austria), Dr. Jairo Patiño (University of Liège, Belgium), Dr. Jesús Muñoz (Royal Botanical Garden of Madrid, Spain), Dr. Juana María González-Mancebo (La Laguna University, Tenerife, Canary Islands, Spain), Dr. Maria Privitera and Dr. Marta Puglisi (Catania University, Italy), Dr. Michele Aleffi (Camerino University, Italy), Mr. Roland Barth (Nature centers of Bellevue, Nebraska, U. S. A.), Dr. Tamás Pócs and the Eszterházy Károly College Herbarium (Hungarian Academy of Sciences, Eger, Hungary) and Mr. Tom Blockeel (Sheffield, United Kingdom), for sharing information and borrowing or collecting plant material which were essential to the completion of this thesis. Special thanks are given to the great and critical help of Dr. Susana Rams (Centro de Magisterio "La Inmaculada", Universidad de Granada, Spain) with Sierra Nevada description.

I am grateful to the Plant Biology Department, Faculty of Biology, International Relationships Service and doctoral program of Biodiversity and Environmental Management, of Murcia University for facilitating my research and stay in Murcia. The Spanish Ministry of Foreign Affairs, Spanish Agency for International Cooperation and Development (Agencia Española de

Cooperación internacional y Desarrollo, AECID), Department of Scientific Cooperation, Scholarships section for the financial support in form of a PhD-Scholarship during the academic years 2009-2010, 2010-2011 and 2011-2012. The Spanish Ministry of Science and Innovation for funding the research projects in which this thesis has been carried out (projects code: CGL2008-00275/BOS and CGL-2011-22936/BOS). The management group of Sierra Nevada Natural Space is kindly acknowledged for their permission to collect samples in the National and Natural Parks.

Special thanks and my dearest sentimental acknowledgement to persons that affected my life by their extreme support and encouragement to be the person I am, ordered by their priceless appearance in my life:

In Egypt, all my respectful gratitude to my family, especially my mother and father (God may bless their souls) who gave me the moral as well as financial support I required during all my life, and to the dearest person to my heart, my wife Samah, for her constant love and support.

In Faculty of Agriculture, Ain Shams University, where I started my career, I deeply thank, Dr. Mohamed Tahaa, Dr. Nabil Killiny, Dr. Mahmoud Yousef, Dr. Mohamed Abdel Sallam Rashid, Dr. Abdel-Fatah Awad, Dr. Fatthy Abdel Tawab, Dr. Samir Abdel Aziz, Dr. Eman Fahmy, Dr. Ashraf Bakry, Dr. Nouh Eid, Dr. Mahmoud Sallam, Dr. Aiman Atta, Dr. Tarek Tantawy, Dr. Sabry Mousa, Dr. Abdallah Abo-elnasr, M.Sc. Akram Anwar, M.Sc. Mohamed Ahmed Mahmoud Abbas, M. Sc. Ahmed Hal, M. Sc. Fawzya Salah, Hala Zougli, Rahma Ali Hassan, Ahmed Garhi (Saudi Arabia) and many more, for their guidance, support and friendship.

In Spain, since my first day I was blessed by countless adorable Spanish friends and others of different nationalities, who shared with me the almost four years I spent in Spain and made my life easier away from home. They contributed in my life in a way that I will never forget neither them nor the memories we behold together:

- Great friends I met outside the university life: Michael (Germany), Shelli Balta (Sweden), Coc Koun (Turkey) Ahmadu (Cameroon), Yuliaxis

(Cuba), Mehdi Eid and Yassine Mohamed (Algeria), Dr. Faleh Ameen, Dr. Mohamed Ragab, Dr. Tarek Saloma (Egypt), Dr. Jesús Delgado (Perú), Antonio del Valle, Manolo and Javier Vaca (Spain), Dr. Rabeh Mohamed (Tunis) Dr. Mohamed Alkasadi (Yemen), Badreddine Mukel (Syria), Camal & Nesreen Kurd (Turkey) and Dr. Ravindran Jaganathan (India).

- Great friends I met through the University: José Ramón, Inma Aledo, Dr. José Sebastián Carrión, Dr. Santiago Fernández Jiménez, Elena Fierro, José Galián, Iara Chapuiz, Marta Nieto, Elena Castro, Saúl Manzano (all of them from Spain), Paula Mendez and Agripina Ramírez (Dominican Republic), Daniel Achel (Ghana), Hammady Rabeh (Mauritania) and Esmat Rayan (Egypt).

I counted all I remembered instantly, while others are in my heart forever even if I forgot to mention.

Thank you All...

TABLE OF CONTENTS

RESUMEN

La especie *Funaria hygrometrica* Hedw. pertenece a la familia Funariaceae, orden Funariales, división Bryophyta. Este musgo se puede identificar fácilmente por los numerosos esporófitos que se entrelazan entre sí, simulando cuerdas, formadas por los pedicelos de las cápsulas que giran en respuesta a los cambios de humedad. Se trata de un musgo cosmopolita con estrategia de vida fugitiva, capaz de adaptarse a diferentes condiciones ecológicas. En ocasiones se desarrolla sobre sustratos quemados. Se caracteriza por su crecimiento rápido y fácil manipulación *in vitro.*

Las poblaciones de *Funaria hygrometrica* de Sierra Nevada, uno de los sistemas montañosos más altos de la Región Mediterránea, fueron estudiadas a lo largo de un gradiente altitudinal que abarcó cuatro de los cinco pisos bioclimáticos presentes en estas montañas (Termo-, Meso-, Supra- y Oro-mediterráneo).

Los objetivos de este trabajo fueron:

1. Estudiar la estructura poblacional y la variabilidad genética de *Funaria hygrometrica* en Sierra Nevada (España).

2. Aumentar el número de datos moleculares conocidos sobre *Funaria hygrometrica.*

3. Establecer nuevos marcadores para el análisis de la variación genética en *Funaria hygrometrica.*

4. Buscar *loci* genéticos relacionados con el proceso de adaptación que *Funaria hygrometrica* sigue en entornos con condiciones climáticas extremas.

5. Testar la posible correlación entre la diversidad genética y las variables ambientales estudiadas, a fin de comprender los factores que limitan la distribución de las poblaciones en Sierra Nevada.

6. Encontrar un escenario hipotético adecuado de la distribución de *Funaria hygrometrica* a lo largo del gradiente altitudinal en Sierra Nevada.

7. Demostrar el potencial de *Funaria hygrometrica* como organismo modelo para comprender y seguir su respuesta y adaptación al cambio climático.

Fueron seleccionados varios marcadores moleculares con el fin de comparar y complementar la información obtenida. Los marcadores escogidos fueron: secuencias de ADN nuclear, cloroplastidial y mitocondrial, así como técnicas de huellas genéticas como microsatélites y AFLP.

El material vegetal se recolectó en las montañas de Sierra Nevada (España) durante dos viajes de muestreo (Mayo de 2010 y Julio 2011). Se prospectaron 17 localidades, en las que se recolectaron un total de 160 muestras, las cuales fueron ordenadas por su altitud sobre el nivel del mar en metros. Finalmente, sólo 84 de ellas pudieron ser utilizadas en los análisis moleculares. Las localidades fueron divididas en cuatro grupos de acuerdo con la altitud donde estaban ubicadas. El grupo uno (G1): poblaciones de zonas bajas (situadas por debajo de 600 m), grupo dos (G2): poblaciones de altitud media (entre 600 m y 1 000 m), grupo tres (G3): poblaciones de altitud media-alta (entre 1 000 m y 1 800 m) y el grupo cuatro (G4): poblaciones de zonas altas (entre 1 800 m y 2 700 m). Además, se obtuvieron 15 muestras de *Funaria hygrometrica* de la cuenca Mediterránea (Grecia, Italia y Turquía), Europa occidental (Reino Unido: Inglaterra), Norteamérica (EE.UU: Carolina del Norte, California y Connecticut) y Asia (Rusia: Siberia occidental). Todas las muestras fueron identificadas y confirmadas en el laboratorio. Se llevaron a cabo cultivos *in vitro* de todas ellas con el fin de estar seguros de que la diversidad genética obtenida correspondía a un esporófito y se eliminaba así la posibilidad de contaminación por microorganismos naturales del suelo. Los cultivos se iniciaron a partir de esporas provenientes de cápsulas maduras, excepto en el caso de las muestras de las localidades 14, 15, 16 y 17 ya que las cápsulas de estas muestras contenían esporas no viables y por tanto se utilizaron los gametófitos verdes esterilizados para la extracción directa del

ADN. Las muestras fueron cultivadas en un medio nutritivo Murashige y Skoog (MS) o un medio sólido Knop (KM). Todos los cultivos se mantuvieron en una incubadora a 22 ± 3 ºC y con 16/8 h de luz/oscuridad suministrada por tubos fluorescentes blancos a un flujo de fotones de 33,5 a 55,0 µmol/sm^2. Los protonemas obtenidos con éxito se utilizaron para la extracción del ADN.

La amplificación del ADN a través de la reacción PCR fue realizada utilizando inicialmente varias regiones como candidatas. Entre las regiones del ADN nuclear, fueron probadas la región ITS y el gen *FLORICAULA/LEAFY* homólogo 1 (*PpLFY*1). Para la región ITS se utilizaron los cebadores propuestos por Douzery *et al.* (1999) con algunas modificaciones, introducidas con el fin de que coincidieran con la secuencia de *Funaria hygrometrica* publicada en GenBank (X74114). Para el gen *FLORICAULA/LEAFY* se rediseñaron los cebadores de Tanahashi *et al.* (2005) basándose en el genoma del musgo *Physcomitrella patens*. Para la amplificación de ADN cloroplastidial, inicialmente se pusieron a prueba los cebadores del intrón *trn*V (Werner *et al.* 2009), los de la región *trn*K$_{UUU}$ (con el gen *mat*K gen incluido, Borsch y Quandt 2009), los de la región *pet*N-*pet*D (Liu *et al.* 2012a) y los del gen de la ARN polimerasa Rpb1 (*rpo*C1, Royal Botanic Gardens of Kew 2012). En el caso de la región cloroplástica *rps3-rpl*16 (que incluye el espaciador *rps*3-*rpl16*, el exón I del gen *rpl*16 y el intrón del gen *rpl*16) se diseñó un conjunto completamente nuevo de cebadores con el fin de mejorar la calidad de las secuencias. Para el ADN mitocondrial se probaron los cebadores de los espaciadores intergénicos *rpl5-rpl*16, *atp1-trn*W, *atp7-atp*6 y el intrón *rpl*2 (Liu *et al.* 2012b).

Finalmente cuatro regiones fueron seleccionadas para secuenciar el conjunto de muestras: los espaciadores nucleares ITS1 e ITS2, la región cloroplástidial *rps*3-*rpl*16 y el espaciador mitocondrial *rpl*5-*rpl*16. Todas las secuencias fueron verificadas y alineadas.

Para cada región del ADN se analizaron las secuencias obtenidas con el fin de estimar la variabilidad genética y el polimorfismo del ADN y de generar árboles filogenéticos a nivel haplotípico (es decir, secuencias idénticas en las muestras) y a nivel de muestras (de cada muestra

representada por una secuencia). Se calcularon varios índices de variabilidad genética, como el número de haplotipos por localidad (h/loc), la diversidad de haplotipos (H_d) y la diversidad de nucleótidos (π %), los sitios con inserciones/deleciones y las frecuencias de haplotipos. Se llevó a cabo un test de desequilibrio de ligamiento de las secuencias con el fin de probar la asociación no aleatoria entre las regiones secuenciadas. Se estimó el flujo génico (N_m) y se realizaron matrices de distancia genética por pares. Estas últimas se utilizaron para testar la diferenciación de la población mediante el cálculo de la varianza molecular (AMOVA). Los análisis filogenéticos se realizaron utilizando métodos descriptivos no espaciales. Sólo los sitios variables se utilizaron para la reconstrucción filogenética mediante la exclusión de los sitios invariables en todas las secuencias. Se infirieron árboles de expansión mínima construidos a nivel de haplotipo. También se realizó un análisis neighbor joining (NJ), basado en la distancia genética (para el caso en el que se consideraron las estructuras secundarias) con el fin de mejorar la señal filogenética a nivel de haplotipo, utilizando para ello el kit de herramientas disponible en el servidor de la base de datos del ITS2. Los árboles construidos a nivel de muestra se infirieron utilizando tres métodos filogenéticos basados en las secuencias: parsimonia máxima (MP), máxima verosimilitud (ML) e inferencia Bayesiana (BI). Cuando fue posible, la construcción de los árboles se realizó utilizando una especie próxima como grupo externo. Siempre que fue necesario se llevó a cabo un análisis de predicción de estructuras secundarias para confirmar la importancia de la pérdida de fragmentos grandes.

Basándose en la información cedida por el Dr. Goffinet sobre secuencias simples repetidas (SSR) en *Funaria hygrometrica* (también conocidas como microsatélites), fueron probados 20 *loci*, de los cuales fueron seleccionados cinco sobre la base del nivel de polimorfismo y de la consistencia de amplificación en todas las muestras. Éstos fueron amplificados en reacciones Múltiples en condiciones estándar y etiquetados con fluorescencia para separarlos en un secuenciador automático. Como marcadores codominantes, los *loci* SSR se leen de forma alélica por todos los programas especializados. Se probó el grado de idoneidad de todos los *loci*

para calcular la diversidad genética. El polimorfismo genético fue estimado a través de la medición de diversos índices como la diversidad genética no sesgada (UH_Z), la heterocigosidad media observada (OH_z), el número medio de alelos/*locus* (n_A), el número efectivo de alelos (n_e) y el índice de contenido polimórfico (PIC). La estructura de la población se determinó a través del método de agrupación Bayesiana no espacial. Se estimó la diferenciación genética mediante el análisis de la varianza molecular (AMOVA), la distancia genética por pares basada en F_{ST} y el flujo génico.

La técnica de polimorfismos en la longitud de fragmentos amplificados (AFLP) se utilizó de acuerdo con el protocolo original de Vos *et al.* (1995), pero con pequeñas modificaciones. Un total de 12 combinaciones diferentes (Eco + 3 pb x Mse + 3 pb) se marcaron con fluorescencia y se amplificaron para analizar los fragmentos obtenidos en un aparato de secuenciación automática. Con el fin de minimizar la influencia de factores como bandas pequeñas, “bandas tartamudeo”, dímeros y “fondo ruidoso”, a los que el tratamiento de los datos AFLP es muy sensible, se usaron los programas PEAKSCANNER (que analiza los picos fluorescentes) y RAWGENO (para la asignación automática de bandas).

El análisis de los datos AFLP se realizó de dos maneras. Una de ellas basada en el criterio banda-binario (codificando la presencia de bandas detectadas con 1 y su ausencia con 0) y la otra considerando la frecuencia de los alelos (número de veces que está presente una banda en relación al número total de individuos). Cada una tiene sus ventajas y desventajas, por lo tanto, se combinaron ambas formas para obtener el número máximo de índices útiles. Se probó el grado de idoneidad de todos los *loci* para calcular la diversidad genética. El polimorfismo genético se estimó a través de la medición de los siguientes índices: porcentaje de *loci* polimórficos (PPL), número de bandas polimórficas (PB), número bandas polimórficas al 5 % (PB al 5 %), proporción de *loci* polimórficos al 5 % expresado como un porcentaje (PLP al 5 %), número de bandas privadas (PRB), número de bandas fijas privadas (FB), número efectivo de alelos (n_e) y la heterocigosidad esperada bajo la hipótesis de Hardy-Weinberg (H_e). La estructura de la población se determinó a través del método de agrupación Bayesiana noespacial. Se

estimó la diferenciación genética mediante el análisis de la varianza molecular (AMOVA), la distancia genética por pares basada en F_{ST} y el flujo génico. Finalmente, se generó una red de haplotipos.

Las variables bioclimáticas y climáticas fueron extraídas de WorldClim. Ésta página web proporciona un conjunto de 19 variables bioclimáticas que se derivan de la temperatura mensual y los valores de precipitación con el fin de generar variables biológicamente significativas. Las variables bioclimáticas representan las tendencias anuales (como la temperatura media anual y la precipitación anual), la estacionalidad (como el rango anual de temperatura y de precipitación) y las condiciones extremas o limitantes (como la temperatura del mes más frío y del más cálido y la precipitación de los cuatrimestres húmedos y secos). Los valores numéricos de cada variable bioclimática representan el promedio a lo largo de un año. Las variables climáticas son promedios mensuales de temperatura máxima, temperatura mínima y precipitación. Como el ciclo de la vida de *Funaria hygrometrica* se completa en aproximadamente 4 meses y las muestras fueron recolectadas en mayo y principios de julio, sólo fueron considerados los datos de febrero a junio. Se estudió la relación entre la variabilidad genética que se encuentra a nivel de localidad y otros factores asociados a cada una de ellas como las variaciones climáticas, la distancia geográfica o el gradiente altitudinal. Se utilizaron dos métodos para encontrar y medir tales relaciones: 1) la exploración del genoma para localizar los *loci* bajo selección (atípicos) entre los cuatro grupos altitudinales propuestos, y 2) pruebas de correlación entre la estructura genética de la población y, a) la distribución espacial de la variación climática y bioclimática (también conocida como aislamiento por adaptación), b) la distancia geográfica (también conocida como aislamiento por distancia). Las pruebas simples de Mantel (para medir la asociación entre dos matrices) se realizaron para testar todas las combinaciones posibles. Se utilizaron las matrices de la distancia genética y las del logaritmo de la distancia genética de cada marcador (incluyendo los 12 AFLP *loci* atípicos) frente a las matrices de la distancia Euclidiana y del logaritmo de la distancia Euclidiana de las variables ambientales (geográficas, climáticas y bioclimáticas) para encontrar las coeficientes de correlación más altos.

Las cuatro regiones secuenciadas mostraron de manera constante dos grupos de diferente longitud. Los haplotipos fueron generados considerando tanto los sitios polimórficos como los *indels*. El número de haplotipos varió de 7 a 25, dependiendo del *locus*. La diversidad genética detectada para cada *locus* secuenciado varió entre 0,37 y 0,72. Sobre la base del valor de π, el ADN cloroplastidial y el mitocondrial mostraron menos variabilidad que las regiones ITS1 e ITS2. La prueba de neutralidad de Tajima para ITS2 mostró que era neutral (aunque tenía un valor D negativo), mientras que ITS1 y los ADN cloroplastidial y mitocondrial no eran neutrales (valor D significativamente negativo). Sobre la base de los resultados de los árboles de expansión mínima, las muestras parecían formar dos grupos (A, el más frecuente y B, el menos frecuente). Se detectó un tercer grupo (C) en los casos del ITS1, ITS2 y ADN cloroplastidial, aunque generalmente estaba relacionado con el grupo A. Los árboles filogenéticos generados a nivel de muestra con los tres métodos utilizados, confirmaron estos resultados con un apoyo elevado en los casos de ITS1, ITS2 y ADN cloroplastidial. El análisis de la estructura secundaria ayudó a detectar la importancia de la pérdida del fragmento grande en el espaciador intergénico *rpl*15-*rpl*16 del ADN mitocondrial, lo que puede ser considerado como un solo evento mutacional. Mientras que en la región cloroplástica *rps*3-*rpl*16 mostró una menor posibilidad de ser un evento mutacional único. Se utilizó la predicción de la estructura secundaria para mejorar la señal filogenética con el kit de herramientas disponible en el servidor de la base de datos del ITS2, que utiliza el método NJ para encontrar el mejor árbol. Las estructuras secundarias del grupo C resultaron, con un elevado apoyo bootstrap, estar más relacionadas con el grupo A que con el grupo B.

Las técnicas de huellas genéticas fueron polimórficas. La diversidad genética no sesgada de los microsatélites (UH_Z) fue de 0,45, mientras que el valor del índice de contenido polimórfico (PIC) fue de 0,41. Esto refleja un nivel de polimorfismo moderado entre las localidades muestreadas en comparación con los otros marcadores. En el análisis AFLP se obtuvo con éxito un número adecuado de *loci* (1 584), con una heterocigosidad media esperada (H_e) de 0,21, que refleja la diversidad más baja detectada entre los

marcadores genéticos utilizados. El análisis de los *loci* bajo selección (atípicos) reveló que 12 *loci* parecen estar bajo la selección positiva entre los cuatro grupos altitudinales propuestos, que corresponden aproximadamente a los pisos bioclimáticos de Sierra Nevada. Sobre la base de los conocimientos actuales, este estudio parece ser el primero en usar el análisis de detectar *loci* candidatos sometidos a selección basado en la exploración del genoma del musgo *Funaria hygrometrica* y podría ser incluso el primero reportado en todos los musgos.

Tras los análisis filogenéticos realizados con las secuencias obtenidas y los análisis de la estructura de las poblaciones con los datos de las técnicas de huellas genéticas, se observó que las muestras que aparecen agrupadas proceden de localidades diferentes. El flujo génico estimado a partir del F_{ST} y su análogo Φ_{ST} para ITS1, ITS2, ADN cloroplastidial, ADN mitocondrial, SSR, AFLP y 12-AFLP fueron 0,85, 0,44, 0,52, 0,27, 2,00, 1,88 y 0,30 respectivamente. Los 12 *loci* AFLP atípicos detectados mostraron niveles elevados de diferenciación genética, lo que indica que están sometidos a selección. Los valores de diferenciación genética para ITS1, ITS2, ADN cloroplastidial y ADN mitocondrial también fueron relativamente altos para un organismo con una elevada capacidad de dispersión a distancias relativamente cortas.

La media de la temperaturas máximas durante el ciclo de vida de *Funaria hygrometrica* en Sierra Nevada se correlacionó significativamente (valor de $p \leq 0,01$) con el ITS1 y todo el conjunto de *loci* AFLP (valor de $r = 0,52$ y 0,51, respectivamente). Mientras que la media de las temperaturas mínimas se correlacionó significativamente (valor de $p \leq 0,01$) con los *loci* SSR (valor de $r = 0,43$). La interacción entre los períodos de sequía y la elevada temperatura durante un año completo (la precipitación del mes más seco, la precipitación del cuatrimestre más seco y la precipitación del cuatrimestre más cálido) se correlacionó significativamente (valor de $p \leq 0,01$) con los 12 *loci* AFLP atípicos detectados y los datos de la región ADN mitocondrial. El ADN cloroplastidial y el ITS2 han reflejado niveles insignificantes de correlación con cualquiera de los factores ambientales

probados (valor de p > 0,05). Adicionalmente, ambas regiones resultaron estar significativamente relacionadas (valor de r = 0,64).

En el presente estudio se detectó con éxito y se midió una inesperada elevada variabilidad genética en el musgo *Funaria hygrometrica* a lo largo de un gradiente altitudinal en las montañas de Sierra Nevada, a pesar del tamaño relativamente pequeño de la zona de muestreo. Esta elevada diversidad genética observada en muchas localidades indica que las localidades fueron colonizadas de forma independiente en varias ocasiones. Las colonizaciones múltiples y la estrategia de vida fugitiva de *Funaria hygrometrica* pueden actuar como aceleradores de la evolución adaptativa. Sobre la base de los datos de secuencias se detecta una elevada diferenciación genética entre las localidades muestreadas, mientras que los datos de los marcadores SSR y AFLP muestran cierto grado de homogeneización. No obstante, los 12 *loci* AFLP atípicos fueron los que mostraron la mayor diferenciación genética de todos los marcadores estudiados. Sin embargo, la detección de flujo génico muestra que la diferenciación genética entre las localidades no es completa, lo que rechaza un aislamiento reproductivo entre las localidades muestreadas en las montañas de Sierra Nevada.

Los test de correlación basados en la distancia genética generada a partir de los seis marcadores moleculares sugieren que *Funaria hygrometrica* posee la capacidad de adaptarse genéticamente al gradiente espacial de las medias de temperatura máxima, temperatura mínima y la interacción entre la temperatura y la precipitación. Por consiguiente, *Funaria hygrometrica* sería capaz de adaptarse a los cambios climáticos que pueden ocurrir en las localidades estudiadas en el futuro.

La alta diversidad genética que se encuentra en Sierra Nevada podría ser simplemente el resultado de una alta tasa de migración en un paisaje influenciado por factores humanos o bien debido a migración histórica asociada a cambios climáticos pasados. La existencia de individuos con genotipos mixtos entre los dos grupos principales de *Funaria hygrometrica* sugiere que se trata de dos linajes genéticamente divergentes en lugar de

dos especies crípticas. De los dos linajes genéticos detectados en *Funaria hygrometrica,* uno podría estar especialmente adaptado a los hábitats mediterráneos secos. En elevadas altitudes los dos linajes crecen juntos y se cruzan, mientras que en zonas más bajas el tipo supuestamente mediterráneo es mucho más frecuente, excepto en lugares en los que se dan condiciones ambientales particulares que hacen que haya mayor cantidad de agua disponible. Un escenario posible por el que se habría llegado a la situación actual, podría ser a partir de una población ancestral presente en Sierra Nevada que desarrolló adaptaciones *in situ* a las diferentes condiciones de esta área. La selección local posiblemente fue seguida por la migración bajo el efecto de los cambios ambientales pasados a lo largo del gradiente altitudinal de la sierra. Otro escenario se basa en la idea de que la variabilidad genética en Sierra Nevada podría ser la consecuencia de una migración constante de muestras de *Funaria hygrometrica* conteniendo la diversidad genética de la especie a una escala muy amplia. Pero las esporas solo formarán poblaciones viables si se dan las condiciones adecuadas en el medio para que los gametófitos puedan desarrollarse y reproducirse. Bajo este escenario, los diferentes paisajes de Sierra Nevada habrían permitido el establecimiento de linajes genéticamente distintos, también presentes en muchos otros sitios, quizás incluso a escala mundial. Una hipótesis alternativa podría ser postulada, por la cual la variabilidad genética encontrada en *Funaria hygrometrica* no tendría valor adaptativo. Con el fin de decidir cuál de las tres hipótesis podría estar más cercana a la realidad, los datos obtenidos en Sierra Nevada deberían ser comparados con los obtenidos en un paisaje uniforme de tamaño similar (para confirmar el valor adaptativo de la variabilidad) y con más muestras a escala mundial (para excluir o confirmar la teoría de la adaptación local). Los resultados obtenidos en este estudio apoyan la hipótesis de que la variabilidad genética de *Funaria hygrometrica* en Sierra Nevada es debida a una migración constante de muestras representativas de la diversidad genética de la especie a gran escala. Asimismo, soportan la teoría de que los musgos logran amplias distribuciones geográficas en grandes gradientes ecológicos (por ejemplo, los sistemas montañosos) por medio de la variación genética adaptativa.

ABSTRACT

The species *Funaria hygrometrica* Hedw. belongs to the family Funariaceae, order Funariales, division Bryophyta. This moss can be identified easily by the numerous sporophytes that intertwine with each other, as the tangles formed by the pedicels of the capsules twist in response to changes in humidity. It is a cosmopolitan moss with fugitive life strategy, able to adapt to different conditions and might appears at burnt sites. It is characterized by its rapid growth and easy manipulation *in vitro*.

Populations of *Funaria hygrometrica* in the Sierra Nevada Mountains of Spain (one of highest mountains of the Mediterranean Region) were studied along an altitudinal gradient covering four of the five bioclimatic belts present in this mountainous system (Thermo-, Meso-, Supra- and Oro-Mediterranean belts).

The objectives of this work were:

1- To study the population structure and genetic variability of *Funaria hygrometrica* in the Sierra Nevada Mountains (Spain).
2- To increase the molecular data known about *Funaria hygrometrica*.
3- To establish new markers for genetic variation analysis of *Funaria hygrometrica*.
4- To search for possible genetic *loci* related to the adaptation process that *Funaria hygrometrica* follows in environments with extreme climatic conditions.
5- To test the possible correlation between the genetic diversity and environmental variables under study, in order to understand the factors limiting the population distribution in the Sierra Nevada Mountains.
6- To find a proper hypothetical scenario of the distribution of *Funaria hygrometrica* along the altitudinal gradient of Sierra Nevada Mountains.
7- To show the potential of *Funaria hygrometrica* as a model organism to understand and follow the response and adaption to climate change.

To do so, different markers were selected in order to compare and complement the obtained information. The chosen markers included

sequences of nrDNA, cpDNA and mtDNA, and fingerprinting techniques as microsatellites and AFLP.

The plant material was sampled in the Sierra Nevada Mountains during two collection trips (May 2010 and July 2011). A total of 17 locations were prospected, in which a total number of 160 samples were collected and ordered by their altitude above sea level in meters. Finally, only 84 of them could be used for the molecular analyses. The locations were divided into four groups according to the altitude were they were collected. Group one (G1): lowland populations (below 600 m), group two (G2): middle altitude populations (between 600 m and 1 000 m), group three (G3): middle-high mountain populations (between 1 000 m and 1 800 m) and group four (G4): highland mountain populations (between 1 800 m and 2 700 m). Additionally, 15 samples of *Funaria hygrometrica* were obtained from the Mediterranean zone (Greece, Italy and Turkey), Western Europe (United Kingdom: England), North America (U.S.A.: North Carolina, California and Connecticut) and Asia (Russia: west Siberia). All plant samples were identified and confirmed in laboratory. With the aim to be sure that the genetic diversity analyzed in *Funaria hygrometrica* corresponded to one sporophyte and the contamination by soil and natural microorganisms was eliminated, cultivation *in vitro* was carried out. It was initiated from spores coming from mature capsules except for locations 14, 15, 16 and 17 as capsules of the samples of these locations did not have viable spores. Therefore sterilized green gametophytes were used for direct DNA extraction. Samples were cultivated on nutrient media Murashige and Skoog (MS) or solid Knop medium (KM). All cultures were kept in an incubator at 22±3 ºC, and 16/8 h of light to darkness, supplied by cool-white fluorescent tubes at a photon fluence rate of 33.5 to 55 μmol/sm^2. Successfully cultivated protonemata were used for DNA extraction.

DNA amplification by PCR was performed using several DNA regions as candidates. For nuclear DNA regions, initially the ITS region and the *FLORICAULA/LEAFY* homolog 1 gene (*PpLFY*1) were tested. For the ITS region the primers of Douzery *et al.* (1999) were used with modifications introduced in order to match the sequence of *Funaria hygrometrica* as published (GenBank: X74114). For the *FLORICAULA/LEAFY* gene, the

primers of Tanahashi *et al.* (2005) were redesigned based on the *Physcomitrella patens* genome data. For the cpDNA amplification, initially the intron region *trn*V primers (Werner *et al.* 2009), the *trn*K_{UUU} region primers (with *mat*K gene included; Borsch & Quandt 2009), *pet*N-*pet*D (Liu *et al.* 2012a) and the RNA polymerase Rpb1 (*rpo*C1) gene primers (Royal Botanic Gardens of Kew 2012) were tested. In the case of the *rps*3-*rpl*16 region (which includes the *rps*3-*rpl*16 spacer, the *rpl*16 exon I and the *rpl*16 intron), a completely new set of primers was designed in order to improve the sequence quality. In the case of mitochondrial DNA (mtDNA) primers of the intergenic spacers *rpl*5-*rpl*16, *atp*1-*trn*W, *atp*7-*atp*6 and the *rpl*2 intron (Liu *et al.* 2012b) were tested.

Finally four sequenced regions (ITS1, ITS2, cpDNA *rps*3-*rpl*16 region and mtDNA *rpl*5-*rpl*16 spacer) were selected to sequence the complete set of specimens. All sequences were checked and aligned.

For each sequenced region, data were analyzed to estimate genetic variability and DNA polymorphism and to generate phylogenetic trees on haplotype level (i.e. identical sequence across samples) and sample level (i.e. each sample represented by one sequence). Several genetic variability indices were calculated as the number of haplotypes *per* location (h/loc), haplotype diversity (H_d), nucleotide diversity (π %), observed sites with indels (insertions/deletions) and haplotype frequencies. Linkage disequilibrium analyses, which test the non-random association between sequenced regions were performed. Gene flow (N_m) was estimated and pairwise genetic distance matrices were carried out. The latter were used to test the population differentiation by calculating the molecular variance (AMOVA). Phylogenetic analyses were performed using non-spatial descriptive methods. Only variable sites were used for phylogenetic reconstruction by excluding invariable sites among all sequences. Minimum spanning trees were constructed at the haplotype level. Alternatively a distance-based approach, neighbor joining (NJ) was used in case that the secondary structures were considered to improve the phylogenetic signal, using a web-based toolkit (ITS2 database server). Trees constructed at sample level were inferred using three sequence-based phylogenetic analysis approaches: maximum

parsimony (MP), maximum likelihood (ML) and Bayesian inference (BI). Tree construction was performed including a closely related species whenever it was possible to be used as an outgroup. Whenever it was needed, a secondary structure prediction analysis was conducted to confirm the mutational value of losses of large fragments.

Based on the information shared by Dr. Goffinet about simple sequence repeats (SSR) *loci* -also known as microsatellites- in *Funaria hygrometrica*, 20 *loci* were tested of which five were chosen based on polymorphism level and consistent amplification in all samples. The five *loci* were amplified in Multiplex reactions under standard conditions while labeled fluorescently to separate them on an automatic sequencer. As co-dominant markers, SSR *loci* are read in an allelic form by all specialized programs. All *loci* were tested for adequacy to calculate genetic diversity. Genetic polymorphism was estimated through the measurement of various indices as unbiased gene diversity (UH_Z), mean observed heterozygosity (OH_z), mean number of allele/*locus* (n_A), effective number of alleles (n_e) and the polymorphic index content (PIC). Population structure was determined through a non-spatial Bayesian clustering method. Genetic differentiation was estimated through the analysis of molecular variance (AMOVA), F_{ST} pairwise genetic distance and gene flow (N_m).

The amplified fragment length polymorphism (AFLP) technique was used according to the original protocol of Vos *et al.* (1995) with minor modifications. A total of 12 different combinations (Eco + 3 bp x Mse + 3 bp) were fluorescently labeled and amplified to analyze the obtained fragments in an automated sequencer. The treatment of AFLP data is very sensitive to factors like small bands, stutter bands, dimers, and noisy background. To minimize the influence of such factors an automated AFLP scoring was performed using both programs PEAKSCANNER (used for peak analysis) and RAWGENO (for automated scoring).

The analysis of the AFLP data was performed in two known approaches. The first is based on band-binary criterion (i.e. codifying the detected bands to 1 when presence and 0 when absent) and the other is

based on allele frequency (i.e. number of a band presence relatively to the number of all individuals). Each has its advantages and disadvantages, therefore, both forms were combined to obtain the maximum number of valuable indices. All *loci* were tested for adequacy to calculate genetic diversity. The genetic polymorphism was estimated through the measuring of indices like percentage of polymorphic *loci* (PPL), number of polymorphic bands (PB), number of polymorphic bands at 5 % (PB at 5 %), proportion of polymorphic *loci* at the 5 % level - expressed as a percentage - (PLP at 5 %), number of private bands (PrB), number of fixed private bands (FB), effective number of alleles (n_e) and expected heterozygosity under Hardy-Weinberg assumptions (H_e). Population structure was revealed through a non-spatial Bayesian clustering method. Genetic differentiation through the analysis of molecular variance (AMOVA), F_{ST} pairwise genetic distance and gene flow (N_m) were estimated. Finally a haplotype network was generated.

Bioclimatic and climatic variables were taken from WorldClim database. It provides a set of 19 bioclimatic variables, which are derived from the monthly temperature and rainfall values in order to generate biologically meaningful variables. The bioclimatic variables represent annual trends (e.g., mean annual temperature, annual precipitation) seasonality (e.g. annual range in temperature and precipitation) and extreme or limiting environmental factors (e.g. temperature of the coldest and warmest month, and precipitation of the wet and dry quarters). The numerical values of each bioclimatic variable represent the average over a year. Climatic variables include monthly averages of maximum temperature, minimum temperature and precipitation. As *Funaria hygrometrica*'s life cycle is completed in approximately 4 months and the sampled were collected in May and first days of July, only the data from February to June were considered. The relation between the genetic variability found at location level and other factors associated to each location like climatic variations, geographical distance or altitudinal gradient were tested. Two methods to find and measure such relations were used: 1) a genome scan approach to locate the *loci* under selection (outliers) between the proposed four altitudinal groups, and 2) correlation tests between the population genetic structure and, a) spatial patterning of climatic and

bioclimatic variation (a.k.a. isolation by adaptation), b) geographic distance (a.k.a. isolation by distance). Simple Mantel tests (to measure the association between two matrices) were performed by plotting all possible combinations. The genetic distance and log (genetic distance) matrices of each marker (including 12 AFLP positive outlier *loci*) were tested against the Euclidean distance and the log (Euclidean distance) matrices of the environmental variables (geographic, climatic and bioclimatic) in order to find the highest correlation coefficients.

The four sequenced regions resulted constantly in two size groups. Haplotypes were generated based on both polymorphic sites and indels. The haplotype number varied from 7 to 25, depending on the *locus.* The genetic diversity detected for each sequenced *locus* varied from 0.37 – 0.72. Based on π value, cpDNA and mtDNA were more conserved than ITS1 and ITS2. Tajima's neutrality test showed that ITS2 was neutral (although it had a negative D value), while ITS1, cpDNA and mtDNA were not neutral (significantly negative D value). Based on the minimum spanning trees, the samples appeared to be in two groups (A, more frequent and B, less frequent), while sometimes a third group (C) appeared (the case for ITS1, ITS2 and cpDNA), although generally related to group A. Phylogenetic trees supported with three methods on the sample level confirmed these findings. Secondary structure analysis helped to detect the mutational value of the missing large fragment in the *rpl*15-*rpl*16 intergenic spacer of the mtDNA, which can be considered as one mutational event. While in chloroplast *rps*3-*rpl*16 region, uncertainly, it showed a lower possibility for being a single mutational event. Secondary structure prediction was used to improve the phylogenetic signal with a web-based toolkit (ITS2 database server), which uses the NJ method to find the best tree. Group C haplotype's secondary structures were more related to group A rather than to group B supported with high bootstrap values.

Fingerprinting techniques were polymorphic. Microsatellites unbiased gene diversity (UHz) was 0.45, while the value of polymorphic index content (PIC) was 0.41. This reflects a moderate polymorphism level among sampled locations in comparison to the other markers. AFLP was successfully able to

produce an adequate number of *loci* (1 584) for analysis with a mean expected heterozygosity (H_e) of 0.21 which reflects the lowest detected diversity among the used genetic markers. *Loci* under selection (outliers) analysis revealed 12 *loci* to be under positive selection between the proposed four altitudinal groups, which approximately correspond to the bioclimatic belts of Sierra Nevada. To the best of the present knowledge, this study is the first to use the detection of candidate *loci* under selection based on a genome scan in the moss *Funaria hygrometrica* and might be the first report of all mosses.

Following the phylogenetic analyses based on the obtained sequences and the population structure analyses based on the data obtained by the fingerprinting techniques, it appeared that clustered samples are originated from different locations. The gene flow estimated from F_{ST} and its analogue Φ_{ST} for ITS1, ITS2, cpDNA, mtDNA, SSR, AFLP and AFLP-12 were 0.85, 0.44, 0.52, 0.27, 2.00, 1.88 and 0.30 respectively. The detected AFLP 12 outlier *loci* showed high genetic differentiation levels, which is a signature of *loci* under selection. The genetic differentiation values for ITS1, ITS2, cpDNA and mtDNA were also relative high for an organism with a high dispersal capacity at relatively close distances.

During the assumed life cycle of *Funaria hygrometrica* from Sierra Nevada of Spain, the estimated average of maximum temperature was significantly correlated (p-value ≤ 0.01) to the ITS1 and the AFLP all *loci* dataset (r-value = 0.52 and 0.51, respectively). While the estimated average of minimum temperature was significantly correlated (p-value ≤ 0.01) to all SSR *loci* (r-value = 0.43). The interaction between the drought periods and warm temperature over a complete year (the precipitation of driest month, precipitation of driest quarter and precipitation of warmest quarter) was significantly correlated (p-value ≤ 0.01) to the detected 12 AFLP outlier *loci* and the mtDNA sequence data. The cpDNA and ITS2 proved to reflect insignificant low levels of correlation to any of the tested environmental factors (p-value > 0.05), additionally, both were found to be significantly related (r-value of 0.64).

The current study was successfully able to detect and measure an unexpected high genetic variability found for the moss *Funaria hygrometrica* sampled from locations along the altitudinal gradient of the Sierra Nevada Mountains in spite of the relative small size of the sampling area. Such a high genetic diversity in many of the sampled locations indicates that they were colonized independently several times. Multiple introductions and the fugitive life strategy followed by *Funaria hygrometrica* can efficiently accelerate the adaptive evolution. Strong genetic differentiation exists between the locations based on the sequenced regions, while SSR and AFLP markers showed some homogenization. However, the AFLP 12 outlier *loci* revealed the highest genetic differentiation of all studied markers. However, the detection of gene flow shows that the genetic differentiation among locations is not complete, which rejects an absolute reproductive isolation between the sampled locations in Sierra Nevada Mountains.

Correlation tests based on the genetic distance generated from the six molecular markers suggest that *Funaria hygrometrica* possesses the ability to adapt genetically to the spatial gradient of the averages of maximum temperature, minimum temperature and the interaction between temperature and precipitation. Consequently, *Funaria hygrometrica* would be able to adapt to the climate changes that may occur on the sampling locations in the future.

The high genetic diversity found in Sierra Nevada might simply be the result of a high migration rate into a landscape influenced by human factors or due to historical migration associated to past climatic changes. The existence of specimens with mixed genotypes suggests that the two main groups of *Funaria hygrometrica* correspond to two genetically divergent lineages rather than being cryptic species. From the two detected lineages, one might be adapted especially to dry Mediterranean habitats. At higher altitudes the two lineages grow together and cross, while at lower altitudes the supposed Mediterranean type is much more frequent, unless particular environmental conditions make a higher amount of water available. One possible scenario which would explain the current situation could be that an ancestral population present in Sierra Nevada developed *in situ* adaptions to the different conditions of the studied area. The local selection possibly was followed by

migration under the effect of past environmental changes along the altitudinal gradient of the mountains. Another scenario would be that the genetic variability in Sierra Nevada could be the consequence of a constant migration of samples representing the genetic diversity of *Funaria hygrometrica* at a much wider scale. But the spores only will be able to give rise to stable and viable populations if suitable conditions are present in the environment that allow gametophyte's growth and reproduction. Under this scenario, the diverse landscape of Sierra Nevada would permit the establishment of genetically diverse lineages also present at many other sites, perhaps even at a worldwide scale. An alternative hypothesis could be postulated, that the genetic variability found in *Funaria hygrometrica* does not represent any adaptive value. In order to decide which hypothesis is closer to reality, the data obtained in Sierra Nevada should be compared with those obtained from a uniform landscape of areas with a similar size (to confirm the adaptive value of the variability) and with more samples at a worldwide scale (to exclude or confirm the local adaption theory). The presented results favor especially the scenario in which the genetic variability in *Funaria hygrometrica* is due to a constant migration of samples representing the genetic diversity of the species at a wider scale. They also support the theory that mosses achieve wide geographic distributions across large ecological gradients (e.g. mountains) through adaptive genetic variation.

CHAPTER 1

INTRODUCTION

1.1 BIODIVERSITY AND ADAPTATION

Biodiversity is the degree of variation of life forms within a given species, ecosystem, biome, or an entire planet. With more precision, biologists define biodiversity as the "totality of genes, species, and ecosystems of a region" (Larsson 2001; Davis & Cornwell 2006). Four types of diversity levels can be distinguished: ecological diversity, species diversity, genetic diversity and molecular diversity (Campbell 2003). The biodiversity is enriched by the effect of the genetic variability in all types of living organisms. Genetic variability is defined as the measure of the tendency of individuals in a population to vary from one population to another and it is clearly separated from the genetic diversity (Li 2006).

According to the Hardy–Weinberg theorem, both allelic and genotypic frequencies in a population remain constant from one generation to another in the absence of disturbing influences (Hardy 1908). Those disturbing influences are also known as population dynamics: mutations, selection, random genetic drift, gene flow and meiotic drive (Hardy 1908; Crow 1999). The genetic variability is the result of such population dynamics, which is an important factor in the evolution, as it affects an individual's response to environmental changes and thus can lead to differential survival of organisms within a population due to natural selection of the fitter variants. Without variability, it becomes difficult for a population to adapt to environmental changes and therefore it is more prone to extinction (Sousa *et al.* 2011). This adaption process forms part of the evolution of a species, termed "evolvability", which means the ability of a population of organisms to not merely generate genetic diversity, but to generate adaptive genetic diversity, and thereby evolve through natural selection (Altenberg 1995; Kirschner & Gerhart 1998; Colegrave & Collins 2008).

Plant populations can avoid extinction by tolerating environmental changes (e.g. climate) through phenotypic plasticity, adapting to new conditions through selection toward genetic standing variation or favorable mutations, or migrating to locations with favorable conditions (Davis *et al.* 2005; Aitken *et al.* 2008). Adaptive plasticity facilitates the crucial first step in

adaptation to new environments reducing the effect of directional selection (e.g. Haldane 1957) and allowing enough time for a population to become established where standing genetic variation in combination with mutation and/or recombination among individuals can provide a range of heritable phenotypes to respond to local selection pressures (Pigliucci 2001). Such plasticity therefore not only reduces the probability of extinction in new environments, but also allows populations to more easily move from one adaptive stage to another (Robinson & Dukas 1999; Pigliucci & Murren 2003; Price *et al.* 2003; West-Eberhard 2003; Schlichting 2004; Amarillo-Suarez & Fox 2006).

In order to test the adaptive signals in a species, association of genetic variation with environmental variables has been frequently reported in the recent literature (Hedrick *et al.* 1976; Aitken *et al.* 2008; Balkenhol *et al.* 2009). For example, Richardson *et al.* (2009) analyzed climate-related genetic patterns in the western white pine (*Pinus monticola* Douglas ex D. Don) suggesting that divergent climatic selection has influenced phenotypic traits associated with tree growth. Among the climatic variables that contribute to genetic differentiation in plant species, the influence of temperature and changes in precipitation patterns have recurrently been demonstrated since the work of Turesson (1925). An example of the influence of climatic conditions is the timing of phenological events, such as flowering, which depend on temperature and which are obvious targets for natural selection (Stanton & Galen 1997; Stinson 2004; Jump & Peñuelas 2005; Jump *et al.* 2009; Wang *et al.* 2009; Doi *et al.* 2010; Scherrer & Körner 2011).

From this perspective, appropriate standing variation and the ability of long-distance gene dispersal are key factors for plant species to keep track with rapidly changing environments (Davis & Shaw 2001).

1.2 MODEL OF STUDY

1.2.1 Bryophytes

Bryophytes are the only land plants with a dominant, green, photoautrophic, and branched gametophyte, which exhibits a diversity of morphologies unparalleled in tracheophytes (Crum 2001). They are erroneously defined as non-vascular plants, however they are devoid of specialized tissue for the internal transport of water and nutrients (Ligrone *et al.* 2000). The sporophyte is unbranched and permanently attached to the maternal (gametophyte) plant. This sporophyte has a capsule like structure, which yields a single sporangium containing the spores. The first multicellular stage following the germination of the spore is termed a protonema, which is a filamentous or thalloid structure that develops into a gametophore, which holds the gametangia (Sitte *et al.* 1994). In addition, bryophytes are able to propagate by a large variety of vegetative ways such as rhizoidal gemmae, axillary gemmae, brood bodies, detaching leaves or buds, leaf fragments, etc. (Frahm 2008). Although these organs for vegetative propagation are much larger than spores, they have similar effects for the dispersal over large distances (Frahm 2008). The effectiveness of both vegetative and generative propagation is best demonstrated by the re-colonization of Europe and North America after the glaciations in the Pleistocene from refuges in southern latitudes (Frahm 2008).

Bryophytes include mosses, liverworts and hornworts. Their wide range distribution comes mainly from their ability to adapt to the changes in the ecological factors (Vanderpoorten & Goffinet 2009). Bryophytes include about 15 000 known species (Hallingbäck & Hodgetts 2000).

Apart from the economic interests, many fundamental and applicative morphological, genetic, ecological, physiological and evolutionary, as well as other problems can be studied in bryophytes better than in vascular plants, due to certain characteristics that make them suitable as model organisms, environmental monitors and bio-indicators. Some of these characteristics are:

1. Bryophyte gametophytes have almost no buffering capacity against environmental stochasticity. They are more vulnerable than the

sporophytes of seed plants (Proctor 2000). This vulnerability leads to intensive variability in spatial and temporal dynamics in bryophyte populations (Lloret 1994; During & Lloret 1996).

2. High phenotypic plasticity in relation with genetic variation (Mishler 2001) and the low rate of morphological evolution, as fossil evidence suggests that early land plants were structurally similar to extant bryophytes (Kenrick & Crane 1997).
3. Poikilohydric nature and desiccation tolerance. Poikilohydric plants equilibrate their internal water content rapidly to the water potential of the environment, whereas the desiccation tolerance is the ability of a plant to recover after being dried at the cellular level (Mishler 2001). Along with the small size, and the lack of roots, poikilohydry makes bryophytes more dependent to their immediate environment (Mishler 2001). As a consequence they are more sensitive to any environmental changes.
4. Strong persistence of asexual reproduction. Because of the difficulty in many cases to carry out fertilization, some bryophytes have evolved losing the ability of sexual reproduction. Somatic mutations occurring in the gametophyte allow some genetic variation even within clonal organisms (Cove 2000; Mishler 2001).
5. Bryophyte gametophytes are haploid, which allows the immediate *in vitro* identification of mutant phenotypes that are naturally occurring or induced by random insertions, physical and chemical mutagenesis, homologs recombination and/or gene targeting techniques (Cove 2000).
6. In nature, the haploidy of the gametophytes leads to a direct exposure to selection. As a consequence, the variability of genes under natural selection may be considerably reduced (Stenøien 1999; Stenøien & Såstad 2001). In such a haploid system, where the sheltering of deleterious alleles is absent, the propagule bank may form a rich pool of adaptive genetic diversity (Hock *et al.* 2008).

Bryophytes are believed to have shared a common ancestor with flowering plants 200 – 400 Mya (Cove 2005). The two lineages diverged early

in land plant evolution as studies comparing the transcriptome of *Physcomitrella patens* (Hedw.) Bruch & Schimp. with that of *Arabidopsis thaliana* (L.) Heynh. show that at least 66 % of *Arabidopsis thaliana* genes have homologues in *Physcomitrella patens* (Nishiyama *et al.* 2003; Cove 2005). Traditionally, the three groups of bryophytes (mosses, liverworts and hornworts) were treated as classes of the division Bryophyta. Molecular data, however, revealed that bryophytes are paraphyletic, with liverworts being sister to all other extant land plants, and mosses sister to a clade of hornworts and vascular plants (Qiu *et al.* 2006). As the three bryophyte groups obviously represent independent lineages in land plant evolution, it is now widely accepted to treat them as separate divisions or phyla, Bryophyta (mosses), Marchantiophyta (liverworts) and Anthocerotophyta (hornworts). In contrast to the latter two groups, mosses are scarcely characterized in terms of morphological-anatomical synapomorphies. At the molecular level, however, all three bryophyte lineages are well circumscribed (e.g. Qiu *et al.* 1998, 2006; Stech *et al.* 2003; Cox *et al.* 2004; Shaw & Renzaglia 2004; Groth-Malonek *et al.* 2005; Forrest *et al.* 2006).

1.2.2 Mosses

Mosses show greater structural diversity than other bryophytes (Schofield 1985), being the second largest of the major land plant lineages as they comprise approximately 12 800 species (Crosby *et al.* 1999). Several features distinguish mosses from other bryophytes according to Schofield (1985):

1- The protonema is generally an extensive and branched filamentous phase of the life cycle.
2- The rizhoids (small rootlike structures) are always multicellular and resemble the protonoma, except that they lack chlorophyll and often have brownish pigmented walls.
3- The protonema sometimes produces gemmae (= rhizoidal gemmae or “tubers”).

4- The gametophore (mature plant or gametangium-bearing stalk) is always leafy, and the leaves are generally radially arranged in more than three rows.
5- Antheridia (male organs) and archegonia (female organs) usually have sterile filaments (paraphyses) intermixed among them.
6- Phyllidia of the gametophore are unistratose for the most part, except at the multistratose midrib (costa). The costa may be single, multiple or even absent.
7- Phyllidial cells are commonly elongate and rarely possess trigones.
8- Phyllidial cells have simple small oil bodies, if any.
9- Phyllidia are rarely lobed.
10- The jacket of the sporangium (capsule) generally has stomata.
11- The outer cells of the sporangium jacket lack transverse barlike thickenings or nodular thickenings.
12- The sporangium jacket is always multistratose.
13- The sporangium usually opens by means of an apical lid (operculum) (Fig. 1.1 A, B).
14- When the operculum falls loose, it usually exposes teeth (peristome teeth) that form a ring around the opening. These teeth are often hygroscopic (Fig. 1.1 B).
15- Within the sporangium there is usually a central mass of sterile tissue (columella) (Fig. 1.1 B).
16- The sporangium is usually elevated on an elongate stalk (seta); the seta is usually wiry and made up mainly of thick-walled cells (Fig. 1.1 A).
17- The calyptra (protective sheath) usually is torn loose from the gametophore by the elongating sporophyte and protects the tip (Fig. 1.1 A).
18- The seta or stalk elongates before the sporangium is differentiated, and is a photosynthetic organ for an extended period during sporophyte development. The presence and shape of the calyptra influence the shape and differentiation of the sporangium.
19- Generally the spores are shed from the sporangium over an extended period.

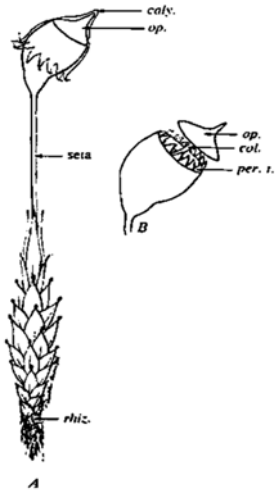

Figure 1.1 Illustration of an idealized moss. A: shows sporophyte (above) emerging from gametophore (below), the phyllidia of which are tipped by gemmae. B: Detail of opening sporangium of A (modified from Schofield 1985).

Mosses in general start their life cycle with a haploid spore cell (n = 1). Whenever suitable conditions are found, the spore (Fig. 1.2 A) starts to germinate in form of protonema (n = 1) through mitotic divisions (Fig. 1.2 B). The protonema then develops to form gametophores (n = 1) (Figs. 1.2 C and D). Gametophores (Figs. 1.2 E and F) produce the antheridia and the archegonia (n = 1). Sperm (n = 1) are released from the antheridia with the assistance of water presence, to reach an archegonium to fertilize an egg (n = 1), which produces a diploid cell (n = 2). This cell through mitotic divisions develops to form a sporophyte. Sporophytes start to appear as small green, pointed orbs tightly arranged in large mats (Fig. 1.2 G). Soon a calyptra shows up (Fig. 1.2 H) and afterwards tangles of stalks with thin, pointed capsules cover the ground in yellowish-green cushions (Fig. 1.2 I). The capsules mature and thicken (Fig. 1.2 J) and they soon drop their calyptra. The stalks and capsules later turn a burnt orange (Fig. 1.2 K) and often remain so through the winter (Fig. 1.2 L). During this time spores are released and they wait for the appropriate conditions to germinate in the next growing season (Parihar 1961).

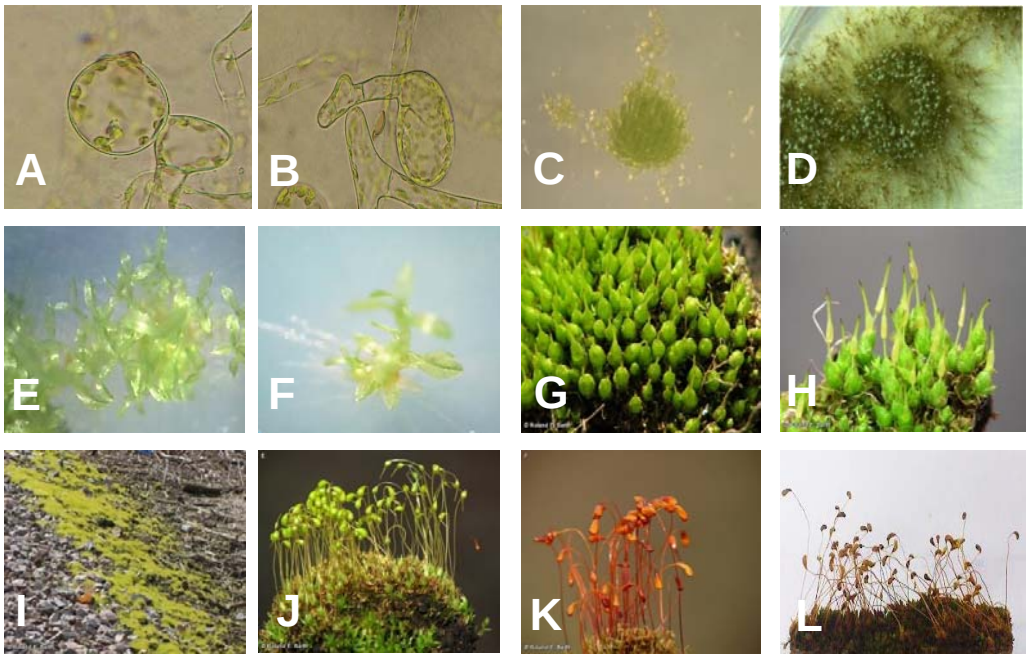

Figure 1.2 *Funaria hygrometrica* in different morphological and physciological states. A to F and L are taken by the author, while the rest are copyrighted photos used after permission of Roland E. Barth (Nature centers of Bellevue, Nebraska)

1.2.3 The common cord moss: *Funaria hygrometrica* Hedw.

Taxon Classification	
Kingdom:	*Plantae*
Phylum:	*Bryophyta*
Class:	*Bryopsida*
Order:	*Funariales*
Family:	*Funariaceae*
Genus:	*Funaria*
Species	*hygrometrica*

Within the Bryophyta, *Funaria hygrometrica* is a member of the family Funariaceae, order Funariales, class Bryopsida. It is one of the easiest to be identified mosses for its peculiar phenotype (sporophytes). It is a weedy moss and grows on moist, shady and damp soil. It can also be found on moist walls, the crevices of rocks and places where recent fires have taken place (Frahm 2008). Its common name is twisted or common cord moss because of the way the numerous sporophytes intertwine with each other, as the tangles formed by the stalks twisting in response to changes in humidity, reflecting its specific epithet (*hygrometrica*). Based on Tropicos (2012) database, there are 36 varieties and two subspecies described in *Funaria hygrometrica*. Most of the varieties probably do not merit the recognition

because of the morphological (phenotypic) plasticity of the species in response to environmental conditions (Miller & Miller 2007). Not even one infra-specific taxon is recognized in Europe (Hill *et al.* 2006), the Mediterranean area (Ros *et al.* 2013), East Europe and North Asia (Ignatov *et al.* 2006), China (Chien *et al.* 2003), Australia (Royal Botanic Gardens Melbourne 2012) and Antarctica (Ochyra *et al.* 2008). Besides the typical form, variety *calvescens* (Schwägr.) Mont. is recognized in North America (Miller & Miller 2007), Mexico (Sharp *et al.* 1994), and Tropical Andean countries (Churchill *et al.* 2000), var. *mauritiana* (Besch.) Paris in the Sub-Saharan Africa (O'Shea 2006), and var. *fuegiana* (Müll. Hal.) Besch. in Chile (Müller 2009).

Funaria hygrometrica is an acrocarpous, autoicous, protandrous and monocarpic moss (which fructify once in the plant life cycle) (Parihar 1961; Crum 1972; Shaw 1991; Hallingbäck & Hodgetts 2000). According to Brugués & Ruiz (2010), Miller & Miller (2007), and Smith (2004) it is described as follows: **Plants** ephemeral, forming yellowish green or green patches, sometimes very extensive. **Rhizoids** light brownish or reddish. **Caulidia** red brownish, 4-30 mm high. **Phyllidia** crowded and bulbiform distally, sometimes laxly foliate throughout, smaller proximally, spreading to erect-patent, imbricate when moist, the upper ones 2-4 mm long, deeply concave, oblong-ovate to broadly obovate, acute to apiculate or short-acuminate, entire or weakly serrulate distally; costa subpercurrent to short-excurrent; distal laminal cells thin-walled and inflated, hexagonal or oblong-hexagonal, 30-50 µm wide, becoming much more oblong proximally. **Seta** usually (12-)20-45(-80) mm long, slender and flexuose, usually hygroscopic. **Capsule** 2.0-3.5 mm long, pyriform, asymmetric, curved to straight, horizontal to pendent or merely inclined or nearly erect, becoming sulcate when dry below the strongly oblique mouth; annulus revoluble, operculum slightly convex; peristome double, exostome brown, papillose-striate proximally and papillose distally, strongly trabeculate, becoming appendiculate distally, forming a lattice by fusion of the tips, endostome segments lanceolate about 2/3 as long as the exostome teeth, yellowish, finely papillose-striate. **Calyptra** cucullate, smooth. **Spores**

mostly 12-23 µm, finely papillose. *Funaria hygrometrica* would take up to 4 months to complete one life cycle *in vitro* (Nakosteen & Hughes 1978).

Funaria hygrometrica is a circumpolar wide-temperate moss (Smith 2004), which is reported worldwide as cosmopolitan (Ochyra *et al.* 2008). It is believed to be ubiquitous and seems to have been distributed by man (Frahm 2008). *Funaria hygrometrica* is widespread in northern and temperate latitudes throughout the Holarctic, and widely scattered but generally infrequent in the tropics and in the temperate zone in the Southern Hemisphere (Ochyra *et al.* 2008). It appears to be absent from the Subantarctic, although it has been discovered on South Georgia (Ochyra *et al*. 2002) and Macquarie Island (Seppelt 2004) but these populations represent recent and temporary introductions, as is the case with those in the Antarctic (Ochyra *et al.* 2008).

In *Funaria hygrometrica*, viability of spores stored at room temperature in a capped jar falls from 100 % to 5 % after 11 years (Vanderpoorten & Goffinet 2009). It is able to produce small sized spores in large numbers that may reach distant sites by many factors (wind, rain, man, insects, etc.). This ability is highly compatible with the fugitive strategy adopted by *Funaria hygrometrica* (During 1979) by colonizing habitats that occur unpredictably and are suitable for growth for a very short period of time such as recently burnt sites (Vanderpoorten & Goffinet 2009). However, the establishment of spores may be hampered by the secretion of substances that prevent the growth of the protonema. Such substances are thought to occur in *Funaria hygrometrica* and appear to be non-specific, preventing all protonema in the vicinity from growing. In such a case, establishment is only possible in the complete absence of the parental gametophytes, which means in newly disturbed areas (Kimmerer 1993; Sundberg & Rydin 2002; Cronberg *et al.* 2006).

In several studies plant physiologists have focused their studies on species like *Funaria hygrometrica*, but also on *Ceratodon purpureus* (Hedw.) Brid., and *Physcomitrella patens,* but it was the possibility to realize crosses *in vitro* that led the latter to be chosen for genetic approaches (Schaefer & Zrÿd

2001). These crosses show direct segregations, so F2 or test crosses are unnecessary (Cove 2005). Like *Funaria hygrometrica*, *Physcomitrella patens* belongs to the Funariaceae and they are therefore closely related (Cove 2000). *Physcomitrella patens* is the first moss to be successfully transformed (Schaefer *et al.* 1991) and has recently been singled-out as the first land-plant, and perhaps more interestingly the first multicellular eukaryote, in which gene targeting occurs with an efficiency similar to that observed in the yeast *Saccharomyces cerevisiae* Meyen ex E. C. Hansen (Schaefer & Zrÿd 1997; Schaefer 2001). The *Physcomitrella patens* genome provides a resource for phylogenetic inferences about gene function and for experimental analysis of plant processes through this plant's unique facility for reverse genetics (Rensing *et al.* 2008) and it is the first completely sequenced genome for mosses (Rensing *et al.* 2002).

For such potentials, *Physcomitrella patens* is frequently used as a comparative model in *Funaria hygrometrica* molecular studies (e.g. Szövényi *et al.* 2010; Rüdinger *et al.* 2011) and both are used in many molecular and phylogenetic studies (e.g. Beckert *et al.* 1999; Liu *et al.* 2012a).

Funaria hygrometrica has been used widely as a model of study or a comparative model with others in phylogenetics, developmental biology, abiotic and/or biotic stress physiology, morphogenesis, growth regulator studies, biochemical synthesis and other aspects of molecular biology (e.g. Bopp 1961, 1976, 1980; Chopra & Rashid 1967; D'Souza & Johri 1999, 2003; Krzesłowska & Woźny 2000; Christianson & Duffy 2002; Giordano *et al.* 2002; Hornschuh *et al.* 2002; Sakakibara 2006; von Schwartzenberg 2006; Liu *et al.* 2012a).

In spite of that, all previous studies that mainly targeted *Funaria hygrometrica* as a model, only contributed to identify its physiological characterization (e.g. Christianson & Hornbuckle 1999; D'Souza & Johri 1999, 2003), with the exception to Shaw (1991), who described the genetic structure and variation of *Funaria hygrometrica* based on morphological and biochemical characters. Two studied populations (mine site and campsite fire pit) differed in all measured traits, but complete monomorphism within

populations at 14 enzyme *loci* suggested that each represented a single clone. Variability in gametophytic growth rates and responses to different experimental media was found, however, it occurred among haploid sib families (families of meiotic progeny derived from the same sporophyte) and among sibs within families within both populations, suggesting high levels of genetic variability that the 14 enzyme *loci* were unable to detect. Low mean reproductive output and a high level of variability among sporophytes in a mine site population probably reflected heavy metal toxicity. Based on this study, in combination with previous work on *Funaria hygrometrica* (Shaw 1990), somatic mutation and/or non-genetic effects appear to contribute significantly to phenotypic variability in natural populations. This study was followed by another publication by Shaw & Bartow (1992), which described *Funaria hygrometrica* that grow in close proximity on contaminated tailings of a copper mine. It displayed very low levels of genetic variability, but had extensive morphological plasticity.

In addition to the previously stated characteristics we used *Funaria hygrometrica* as a model of study in our currently work for the following reasons:

1- The lack of recent information about the genetic variation of this model moss on molecular level.
2- The high growth rate in laboratory conditions in comparison with *Physcomitrella patens* for example (Nakosteen & Hughes 1978).
3- As a cosmopolitan moss, *Funaria hygrometrica* can be found frequently while *Physcometrilla patens* is not a common moss even if it is widely distributed in temperate zones (Cove 2005).
4- The high dispersal ability and fugitive characteristics, which might make this species a suitable model to study its population structure in relation to the changes in environmental factors.

1.3 STUDY AREA: SIERRA NEVADA OF SPAIN

Sierra Nevada (meaning "snowy range" in Spanish) is a mountain range situated in the south of Spain (Fig. 1.3), belonging to the Autonomous Community of Andalusia. It occupies approximately 2 000 km^2, from which about 2/3 belong to the Granada province and 1/3 to the Almeria province. It was declared as a Biosphere Reserve in 1986, while parts of the range were declared as a Natural Park in 1989 and National Park in 1999.

Sierra Nevada is situated about 50 km from the Mediterranean Sea. It is the Spanish mountain range with the highest altitudes in the Iberian Peninsula, having the Mulhacén peak of 3 482 m above sea level (a.s.l.). Many other peaks have more than 3 000 m, as the Veleta (3 392 m), the Alcazaba (3 366 m), Cerro de Los Machos (3 329 m), Puntal de la Caldera (3 225 m), Puntal de Vacares (3 138 m), the Atalaya (3 107 m), Picón de Jeres (3 088 m) and Cerro del Caballo (3 015 m).

Sierra Nevada is part of the Betic Cordillera. The mountainous system as observed today started to be formed during the Tertiary period from the collision of the African and Eurasian continental plates (Channell & Medizza 1981; López-Garrido & Sanz de Galdeano 2000), a process still ongoing.

The materials present in Sierra Nevada are several. The nucleus part is of siliceous nature and very abrupt. The surrounding parts are formed by carbonate materials and are softer in their topography. The peripheral depressions are filled by post-organic sediments (Gómez Ortiz 2002).

During the last ice age, species moved south to escape the colder climate in the north, and as the climate grew warmer again, these species survived by taking refuge in the mountains. The peculiar situation of the Sierra Nevada Mountains in the far south of Europe, allowed it to become a refuge for countless endemisms and improper mid-latitude species (Prieto Fernández 1975).

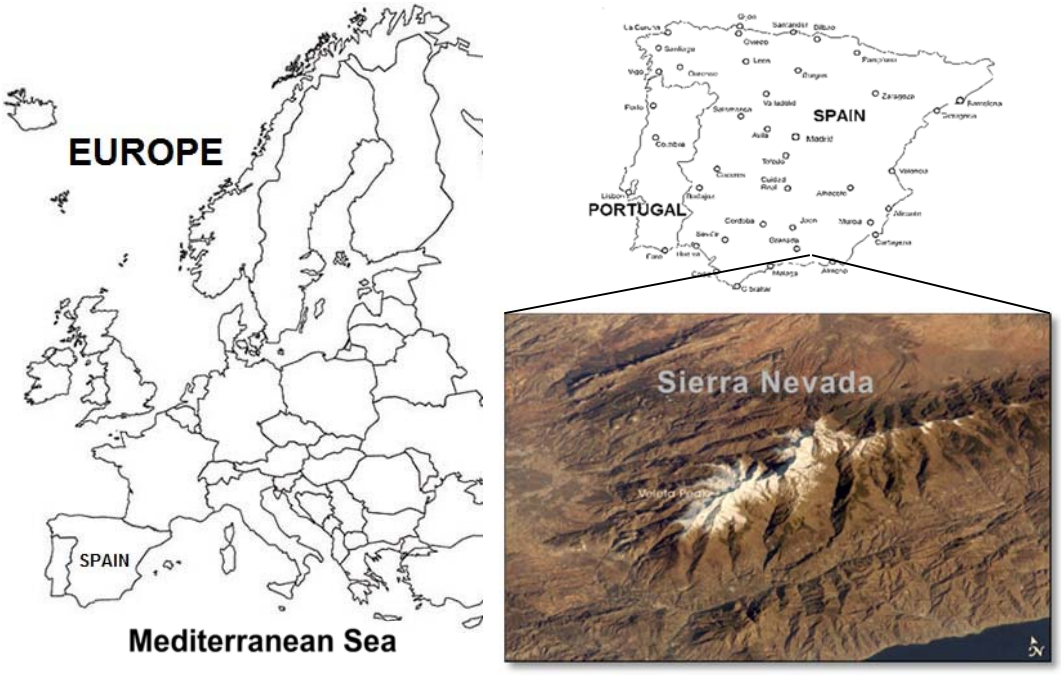

Figure 1.3 Location of Sierra Nevada Mountains (right lower image) in Spain, within the Iberian Peninsula (right upper image), located in European continent (left image).

According to the catalog of national parks of the Ministry of Environment of Spain, Sierra Nevada is the flagship of the Mediterranean high mountain ecosystems with a Mediterranean climate. Therefore, as most characteristic features of the climate of the mountains can be identified by: a) their relative aridity (because of its west-east direction, and the prevailing westerly winds), and b) the strong sunshine of the Mediterranean area.

In Sierra Nevada, five bioclimatic belts or thermotypes can be distinguished (i.e. thermo climatic belt or space which is produced in an altitudinal or latitudinal zonation; Rivas-Martínez 1991). According to their climatic characteristics as described in Molero Mesa *et al*. (1992), starting from base to top they are: Thermo-Mediterranean belt, approximately below 600 to 700 m a.s.l.; Meso-Mediterranean belt, approximately between 800 (600) to 1 300 (1 100) m; Supra-Mediterranean belt, approximately up to 1 800 to 1 950 m; Oro-Mediterranean belt, approximately up to 2 850 (3 000) m and Crioro-Mediterranean belt, higher than 2 850 (3 000) m.

According to Blanca *et al.* (2009), four ombrotypes are present in Sierra Nevada, starting from the dry ombrotype at the base of the sierra, which occupies the major part of the Thermo-Mediterranean belt; the subhumid ombrotype, usually over 1 400 to 1 500 m a.s.l.; the humid ombrotype, between 1 500 to 2 000 m, and the hyperhumid ombrotype over 2 000 m.

The conditions that determine the great diversity in the Sierra Nevada climate are its altitudinal range, meridional latitude in the context of the European continent, and complex topography. The altitudinal range determines the increase of heat stroke on the way up, causing major temperature fluctuations. The north side is colder than the south side for its lower insolation and greater exposure to winds from the north. Its southern position in the Iberian Peninsula and inclusion in the Mediterranean zone, influence its relative dryness in the summer (from May to October) where rainfall is minimal, while in winter rainfall is almost exclusively in the form of snow from a certain altitude of 2 000 m a.s.l. approximately (Blanca *et al.* 2009).

In addition, it should be noted the peculiar microclimatic conditions that are created in each of the valleys, rivers and ravines. The aforementioned sunshine plays an important role in generating more diversity in the mountain flora (Blanca 1991; Díaz 2004; Blanca *et al.* 2009). Many studies aimed this mountain region for its biological richness formed through its peculiar ecological conditions through ages (e.g. Gil 1976, 1988; Molero Mesa *et al*. 1992; Rams *et al.* 2001; Rams Sánchez 2007).

1.4 FUNDAMENTS AND PREVIOUS USE OF THE MOLECULAR TECHNIQUES

Application of molecular techniques and statistical methods of analysis derived from evolutionary theory can be used to estimate how genetic diversity varies spatially and temporally, and to identify crucial forces contributing to present status of genetic diversity (e.g. Moritz 1999; Reed & Frankham 2003; Deyoung & Honeycutt 2005; Whiteley *et al.* 2006). While at the same time, they showed to be helpful to address ecological questions and to find appropriate answers (e.g. Gardner & Latta 2006; Poncet *et al.* 2010; Cooke *et al.* 2012; Manel *et al.* 2012).

Several methods can be used to identify environmental factors determining population structure, for example, through estimates of population divergence (Foll & Gaggiotti 2006). By raising terms like isolation by distance (IBD) and/or isolation by adaptation (IBA), many studies confirmed indirectly the associations between environmental factors (e.g. geographic distance, bioclimatic belts, altitudinal gradient, host-type, etc.) with the genetic divergence, through Mantel statistical test (e.g. Nosil *et al.* 2008; Funk *et al.* 2011).

Alternatively, other approaches try to identify *loci* with clear correlations between allele frequencies and ecological variables and interpret these *loci* as being potentially involved in local adaptation (Joost *et al.* 2007; Coop *et al.* 2010; Poncet *et al.* 2010; Manel *et al.* 2010, 2012).

In bryophytes many studies reported the relation between vegetation diversity and some environmental factors like climate and altitudinal gradient (e.g. van Reenen & Gradstein 1983a, b; Wolf 1993; Frahm & Ohlemüller 2001; Ah–Peng *et al.* 2007; Chantanaorrapint 2010), but without testing the genetic diversity associated to such factors.

1.4.1 Sequencing

Basically, haplotypes are created in the course of evolution via mutations and recombination, and then selection or demographic events such as random genetic drift might occur. Such events are responsible for the spread or disappearance of these haplotypes in populations (Delaneau & Zagury 2012). Standard genotyping (based in PCR/sequencing) allows partial determination of haplotypes depending on the sequenced region. But it is limited because it does not represent a genome in a wide scale as other techniques like AFLP (Gee 2003). The usage of as much as possible sequenced regions and/or genes from different DNA sources would help to obtain more adequate information to define haplotypes and to improve phylogenetic signals (e.g. McMahon & Sanderson 2006; Laenen *et al.* 2011; Liu *et al.* 2012a).

DNA as genetic material can be found in three different places in plant cells (Russell 2001), 1) in the nuclear DNA (nrDNA), 2) in chloroplasts (cpDNA), and 3) in mitochondria (mtDNA or mDNA). Each of them contains genes that are essential for the cell viability and functionality. Normally the genes of nrDNA are structured in a different form than the other two types, as each gene can be divided into untranslated regions (UTRs)-exons-introns, in which introns and UTRs are non-coding regulatory regions in non-overlapped genes, while they can be exons for another different gene in case of overlapped genes (Griffiths *et al.* 2000). Also genes can be found adjacent or separated by space of DNA usually called “spacer DNA” (i.e. non-transcribed DNA between tandemly repeated genes). Its function most likely involves ensuring the high rates of transcription associated with these genes (Russell 2001).

Introns and spacers generally show a lower level of conservation than exons. In introns and spacers mutations accumulate at a higher rate and reflect evolutionary events and genetic variability even between individuals of the same species (Taberlet *et al.* 1991; Manen & Natali 1995). The comparison of their sequences has been useful in phylogenetic, phylodemographic, molecular systematics and molecular ecology analyses in bryophytes (e.g. Werner *et al.* 2007; Laenen *et al.* 2011; Liu *et al.* 2012a).

1.4.1.1 Nuclear DNA

One of the most used polymorphic regions is the Internal Transcribed Spacer (ITS), a space of non-coding RNA situated between structural ribosomal RNAs on a common precursor transcript. ITS spacer is known to be partitioned into ITS1 and ITS2 separated by 5.8S ribosomal cistron, in which the RNA poly-cistronic precursor transcript will be in this order 18S rRNA, ITS1, 5.8S rRNA, ITS2, 28S (Wheeler & Honeycutt 1988). As a part of the transcriptional unit of rDNA, the ITS spacers 1 and 2 are therefore present in all organisms (Calonje *et al.* 2009). Since their first application by Porter & Collins (1991) they are widely used for phylogeny reconstruction, due to the following reasons stated by many early studies (e.g. Baldwin *et al.* 1995; Liston *et al.* 1996; Maggini *et al.* 1998):

1- Biparental inheritance: in comparison to the maternally inherited chloroplast and mitochondrial markers.
2- Easy PCR amplification with several universal primers available for various kinds of organisms.
3- Multi-copy structure, which can be found in up to a few thousand copies *per* cell.
4- Moderate size, which allows reasonable sequencing.
5- Based on published studies they show variation at the level that makes it suitable for evolutionary studies at the species or generic level.

Although they proved to be useful markers, several problems such as flaws in the concerted evolution mechanism, the existence of paralogs and orthologs and the presence of pseudogenes were reported (Mayol & Rosselló 2001; Bailey *et al.* 2003; Nieto Feliner & Roselló 2007; Soltis *et al.* 2008). Consequently, sequences must be carefully analyzed in order to avoid wrong phylogenetic inferences (Mayol & Rosselló 2001; Bailey *et al.* 2003; Nieto Feliner & Roselló 2007). In order to detect these kinds of problems, PCR products should be sequenced using both strands and contigs should be assembled with a high percentage of overlap in order to detect polymorphisms. At present, *Funaria hygrometrica* has only three ITS-

sequences deposited in the GenBank database (X74114.1; X80212.2; JN089174.1), originated from the work of Capesius (1995), Capesius & van de Peer (1997) and Liu *et al.* (2012a), respectively.

The ribosomal DNA gene is not only one of the most important markers in molecular systematics (Álvarez & Wendel 2003; Coleman 2003), it is also used in species barcoding and DNA array technologies (e.g. Engelmann *et al.* 2009; Landis & Gargas 2007). Research of the last 20 years has mainly focused on exploiting the ITS spacer sequences for phylogenetic reconstruction (Schultz & Wolf 2009).

rDNAs encode RNA genes, which are single stranded but develop secondary structure where the molecule folds onto itself to form generally short regions of Watson Crick base pairings (G:C and A:U = T) and the intermediate non-canonical pair (G:U = T) in stems (i.e. paired part in double helix form), and single stranded loops (i.e. unpaired nucleotides separating helices; Gillespie 2004).

The secondary structure of ITS spacers (especially ITS2) is increasingly being used in the field of phylogenetic inference due to several reasons: (i) secondary structure analysis extends the taxonomic applicability of this marker. Whereas the rapidly-evolving ITS spacer is mainly suited for phylogenetic analyses at the species level or lower, adding structural information permits analysis at higher taxonomic ranks (Coleman 2003), (ii) the results from secondary structure analysis can improve the phylogenetic resolution obtained from the primary sequence (Keller *et al.*, 2008), and (iii) stems are generally conserved over evolutionary time and the pairings are maintained by compensatory mutation in which the structural data can add new information via compensatory base change (CBC) analysis and thus provide a tool for species delimitation (Müller *et al.*, 2007) and provide information in terms of optimal base pairing that can aid alignment and the tree reconstruction processes (e.g. Coleman & Mai 1997; Denduangboripant & Cronk 2001; Letsch *et al.* 2010).

1.4.1.2 Extra nuclear DNA

In contrast to the nrDNA, both cpDNA and mtDNA are inherited following a uniparental pathway (i.e. from the mother gamete except for some cases), and both are double helix, circular and small sized DNA in a comparison with the nrDNA (Birky 1995). Introns and spacers present in the extra-nuclear DNA are generally less variable than their nuclear counterparts (Borsch & Quandt 2009) but still show a lower level of conservation than exons. Introns in particular possess a well conserved secondary structure that leads to a mosaic of highly conserved stems and extremely variable parts in between (Michel *et al.* 1989; Cech *et al.* 1994; Cech 1998; Kelchner 2002; Borsch *et al.* 2003), while spacers can be fully, partially or not at all transcribed and contain conserved promoter elements as well as a hairpin structure to end transcription (e.g. Quandt & Stech 2004; Won & Renner 2005).

1.4.1.2.1 Chloroplast DNA

Most higher-plant chloroplast genomes have conserved quadripartite structure (two copies of the inverted repeat and the large and small single-copy regions; Jansen *et al.* 2005). Comparative analysis of the chloroplast genome architecture indicates that the gene order and gene content are highly conserved in most chloroplast genomes (Jansen *et al.* 2005). Many researchers tried to design as universal as possible primers that anneal to conserved genes and thereby span more variable spacers and introns for phylogenetic or barcoding techniques (Borsch & Quandt 2009). Borsch *et al.* (2003, 2005) and Löhne & Borsch (2005) showed that rapidly evolving introns and spacers of the chloroplast genome single-copy regions possess high performance as phylogenetic markers. Therefore, they have been widely used in the phylogenetic analysis field (Borsch *et al.* 2007; Jansen *et al.* 2007; Parks *et al.* 2009; Moore *et al.* 2010).

Many chloroplast non-coding DNA regions (introns or spacers) have been used extensively in mosses, more than in other divisions of bryophytes according to Stech & Quandt (2010). For example, the intron region of *rpl*16

gene, *trn*V and *trn*K_{UUU}, and the intergenic spacer *pet*N-*pet*D and *trn*L-*trn*F have demonstrated a utility in phylogenetic studies in bryophytes (e.g. Borsch & Quandt 2009; Liu *et al.* 2012a). However, other studies used successfully the cpDNA to define population structure and conduct phylogeographic studies, e.g. Powell *et al.* (1995); Hamilton (1999) and Petit *et al.* (2002).

The intergenic spacer *rps*3-*rpl*16 is an unpopular noncoding cpDNA marker, which was reported in a few works, but not in bryophytes (Sakai *et al.* 2003; McMahon & Sanderson 2006; Scarcelli *et al.* 2011) in a comparison with *rpl*16 intron and part of the exon II that are more targeted than the spacer (e.g. Kao *et al.* 1990; Jordan *et al.* 1996; Olsson *et al.* 2009; Löhne *et al.* 2007). A BLAST search (Zhang *et al.* 2000) showed that the only complete nucleotide sequences of the *rps*3-*rpl*16 spacer for mosses in the GenBank database are recorded from the whole cp-genome sequence of *Physcomitrella patens* (GenBank: AP005672.1) and *Syntrichia ruralis* (Hedw.) F. Weber & D. Mohr (GenBank: FJ546412.1).

1.4.1.2.2 Mitochondrial DNA

In land plants, the mitochondrial genome is characterized by a dramatic variation in genome size, gene content, and gene order. Such variation causes a significant evolutionary plasticity in contrast with chloroplast genome (Knoop 2004). Only the structure of the bryophytes mitochondrial genome appears to have been highly conserved during its evolutionary history (Liu *et al.* 2012b).

Unlike nrDNA, which is inherited from both parents and in which genes are rearranged in the process of recombination, there is usually no rearrangements in mtDNA from parent to offspring. However, mtDNA is known to recombine with self-copies within the same mitochondrion. In addition to that, the mtDNA has been used to track the ancestry of many species back hundreds of generations through maternal lineage (Brown *et al.* 1979). Although mtDNA is structurally very variable, it is generally the most slowly evolving of the three plant cell genomes at the sequence level, with

unique exceptions in some plant lineages. The slow sequence evolution and a variable occurrence of introns in plant mtDNA provide an attractive reservoir of phylogenetic information (Knoop 2004).

Mitochondrial *loci* are characterized by low substitution rates and less homoplasy, and are hence suitable to improve phylogenetic inferences at least above the species rank (Qiu *et al.* 2010). However, mtDNA makers are much less commonly targeted than cpDNA and nuclear *loci* for phylogenetic studies in land plants (Olmstead & Palmer 1994; Stech & Quandt 2010). Recently, Liu *et al.* (2012b) suggested that the mtDNA markers they developed should be useful in phylogenetic studies within Funariaceae and likely in other moss lineages because the mitochondrial genome contains independent information in comparison to the chloroplast and nuclear genomes.

Many mitochondrial exons, introns and spacers, have been used in phylogenetic analysis and evolutionary studies in bryophytes (Stech & Quandt 2010; Knoop 2010), for example, *rpl*5-*rpl*16, *atp*1-*trn*W, *atp*7-*atp*6 intragenic introns and *rpl*2 gene. All these regions have shown high level of genetic variation in the family Funariaceae in a comparison with other mitochondrial regions (Liu *et al.* 2012a, b).

Apart from the phylogenetics, mtDNA proved to be useful to define population structure, haplotypes within the same species, and showed an association to geographic distributions in animals and plants (e.g. Olson & McCauley 2002; Dickerson *et al.* 2010; Luzhang *et al.* 2010).

1.4.2 Fingerprinting techniques

1.4.2.1 Microsatellites

The genomes of higher organisms (e.g. plants) contain three types of multiple copies of simple repetitive DNA sequences (satellite DNAs, minisatellites, and microsatellites) arranged in arrays of vastly differing size (Armour *et al.* 1999; Hancock 1999). Microsatellites (defined by Litt & Luty 1989), also known by

other names as: simple sequence repeats, SSRs (Tautz *et al.* 1986), short tandem repeats, STRs or simple sequence length polymorphisms, SSLPs (McDonald & Potts 1997), are the smallest class of simple repetitive DNA sequences.

Some authors define microsatellites as two to eight bp repeats (e.g. Armour *et al.* 1999), others define it as one to six bp repeats (e.g. Goldstein & Pollock 1997), or even one to five bp repeats (Schlotterer 1998). Chambers & MacAvoy (2000) suggested following a strict definition of two to six bp repeats, in line with the descriptions of the original authors.

It is now well established that the predominant mutation mechanism in microsatellite tracts is 'slipped-strand mispairing' (Levinson & Gutman 1987). When slipped-strand mispairing occurs within a microsatellite array during DNA synthesis, it can result in the gain or loss of one, or more, repeat units depending on whether the newly synthesized DNA chain loops out or the template chain loops out, respectively (Eisen 1999).

SSR allelic differences are, therefore, the results of variable numbers of repeat units within the microsatellite structure. One common example of a microsatellite is a dinucleotide repeat $(CA)_n$, where n refers to the total number of repeats that ranges between ten and 100. These markers often present high levels of inter- and intra-specific polymorphism, particularly when tandem repeats number is ten or greater (Queller *et al.* 1993).

Microsatellites have become a powerful tool for research in many fields, like population genetics, conservation biology, etc. (Jarne & Lagoda 1996), due to their high genetic variability (e.g. Bowcock *et al.* 1994, Deka *et al.* 1995) and meanwhile they show a different evolutionary aspect than other molecular markers. Therefore, it is often preferred over other molecular techniques like restriction fragment length polymorphism (RFLP) and random amplified polymorphic DNA (RAPD; Goldstein & Pollock 1997). Microsatellites can be used in most applications of population genetics, from identifying relatives to inferring demographic parameters (e.g. Bowcock *et al.* 1994, Goldstein *et al.* 1996). In bryophytes, for example, Clarke *et al.* (2009) showed that microsatellites revealed greater genetic diversity in Antarctic

Ceratodon purpureus populations than with allozymes, and suggested that the extraordinarily high levels of variation reported previously for RAPD studies were artificially elevated by the presence of contaminants.

Although microsatellite markers have been well developed in many kinds of organisms (Zane *et al.* 2002), the number of markers available for bryophytes remain poor (van der Velde *et al.* 2000; Leonardía *et al.* 2006; Provan & Wilson 2007; Hutsemékers *et al.* 2008, 2012). Many recent studies aimed to increase the SSR loci information available for bryophytes (e.g. Leonardía *et al.* 2006; Provan & Wilson 2007; Liu *et al.* 2010; Goffinet, unpublished data).

1.4.2.2 Amplified Fragment Length Polymorphism (AFLP)

The Amplified Fragment Length Polymorphism (AFLP) technique is increasingly used in phylogenetic and population genomics studies, particularly in non-model organisms for which no prior DNA sequence information is available (Vos *et al.* 1995). This useful and relatively cheap technique is based on complete endonuclease restriction digestion of total genomic DNA followed by selective PCR amplification (using primers with one to four selective bases) and electrophoresis of a subset of reduced-by-selectivity fragments, resulting in a unique, reproducible fingerprint for each individual. The technique targets dominant bi-allelic *loci* (coded 0/1) in a 50–500 bp range. Although the AFLP technique is able to generate a large number of informative markers, the success of this method is compromised by different factors (Pompanon *et al.* 2005) as:

1- The AFLP protocol has to be carefully chosen depending on the species under study. The initial AFLP protocol described for plants by Vos *et al.* (1995) has been already successfully modified for the study of other organisms like vertebrates (Bonin *et al.* 2005) or insects (Paris *et al.* 2010).
2- The quality control of the AFLP procedures like contamination (ensure the usage of sterile material or *in vitro* axenic cultures), peak

homoplasy, reliability of the method or genotyping errors (Bonin *et al.* 2004; Pompanon *et al.* 2005; Paris *et al.* 2010).

3- Negative controls and replicated samples should be included in most experiments (Conord *et al.* 2006; Mráz *et al.* 2007; Karrenberg & Favre 2008; Nosil *et al.* 2008; Meyer *et al.* 2009; Fink *et al.* 2010; Poncet *et al.* 2010).

4- Usage of marker selection algorithms developed to optimize the challenging step of AFLP marker scoring by discarding biases due to subjective and unreliable personal procedure (Herrmann *et al.* 2010; Arrigo *et al.* 2009, 2012), for instance, several improvements in automatic scoring have been proposed and implemented in commercial software. However the recently developed freeware RAWGENO V2, built around an optimized and less time-consuming algorithm. With this software a great time saving and higher productivity can be achieved (Arrigo *et al.* 2012).

Recently, AFLPs have become the marker of choice in population genomics studies aiming to detect adaptive genes in non-model species (Pompanon & Bonin 2012). Wide multi-*locus* screening (also known as genome scan) of *locus*-specific signatures of adaptive divergence was used to identify outlier *loci* responsible for such divergence (Black *et al.* 2001; Luikart *et al.* 2003; Storz 2005). Those outlier *loci* (i.e. *loci* revealed by unusually high level of population differentiation), that are involved in adaptation to local environmental conditions are indeed expected to exhibit increased differentiation among populations (along with a decreased diversity within-population). This is so, because divergent selection favors different alleles in different populations. Increased differentiation may also result in hitchhiking with locally adapted variant at linked sites (Maynard Smith & Haigh 1974), at a rate that depends upon the relative strength of selection and recombination. Population genomics is a powerful approach to understand mechanism of selection and species adaptation at the genomic level. AFLP-based genome scans have been used successfully to detect outlier *loci* linked to adaptation to altitude (Bonin *et al.* 2006; Poncet *et al.* 2010), to host plants (Conord *et al.* 2006; Egan *et al.* 2008; Nosil *et al.* 2008; Manel *et al.* 2010), floral divergence

(Herrera & Bazaga 2008), insecticide resistance (Paris *et al.* 2010), adaptation to soil type (Meyer *et al.* 2009), domestication (Rossi *et al.* 2009), or ecotype divergence (Wilding *et al.* 2001; Campbell & Bernatchez 2004; Savolainen *et al.* 2006).

Most studies using population genomics approaches conclude that a substantial proportion of the genomes analyzed show potential signatures of selection (about 5 % of the analyzed *loci*; reviewed in Nosil *et al.* 2009). The detection of outlier *loci* represents only a first step in a comprehensive understanding of how selection shapes genomic divergence between populations. The further step is to identify exactly the genomic regions or genes and ultimately, the mutations involved in the adaptation. But the lack of sequence information given by anonymous AFLP *loci* remains a major limitation to undertake this second step. As a solution, recently developed techniques based on Next Generation Sequencing (NGS) like Restriction-Site Associated DNA (RAD; Baird *et al.* 2008), allow valuable insights toward a clear and precise identification of the *locus* or region under selection in genome scan methods (for more details see Hohenlohe *et al.* 2012).

1.5 OBJECTIVES OF THE STUDY

As far as we know this is the first work planned to describe the genetic variability of *Funaria hygrometrica* on the DNA level. The study objectives were planned and summarized in the following points:

1- To study the population structure and genetic variability of *Funaria hygrometrica* in the Sierra Nevada Mountains of Spain.
2- To increase the molecular data known about *Funaria hygrometrica*.
3- To establish new markers for genetic variation analysis of *Funaria hygrometrica*.
4- To search for possible genetic *loci* related to the adaptation process that *Funaria hygrometrica* follows in environments with extreme climatic conditions.
5- To test the possible correlation between the genetic diversity and environmental variables under study, in order to understand the factors limiting the population distribution in the Sierra Nevada Mountains.
6- To find a proper hypothetical scenario of the distribution of *Funaria hygrometrica* along the altitudinal gradient of Sierra Nevada Mountains.
7- To show the potential of *Funaria hygrometrica* as a model organism to understand and follow the response and adaption to climate change.

CHAPTER 2

MATERIAL AND METHODS

PART 2.1

MATERIAL

2.1.1 MATERIAL SAMPLING

The plant material was sampled in the Sierra Nevada Mountains of Granada, Spain, and the surrounding environment. The sampling was done during two collection trips. The first one was on the 2nd and 3rd of May 2010, and the second one was on the 8th of July 2011. The sampling strategy was based on the following criteria:

1- Accessible locations. Most locations were situated in the proximities of the main roads and paths of Sierra Nevada.
2- To prevent progeny effect. All samples were collected individually at least 1 m apart from others of the same location.
3- To represent the possible maximum altitude range in a gradient form. Sampling started in the Mediterranean coast, at 50 km from the mountains base up to the mountain crests.

The position of the location was determined by GPS (Global Positioning System) using a specialized GPS-device. A short description of each location including altitude and ecological characteristics was annotated. All relevant data were introduced in a spreadsheet file.

2.1.2 SAMPLING LOCATIONS

A total of 17 locations with a total number of 160 samples were collected. Finally, only 84 of them could be used for the molecular analyses. Locations ordered by their altitude in meters, along with their description are given in table 2.1 and were bookmarked on a geographical map (Fig. 2.1).

The populations were divided into four groups: group one (G1): lowland populations (below 600 m), group two (G2): middle altitude populations (between 600 m and 1 000 m), group three (G3): middle-high mountain populations (between 1 000 m and 1 800 m) and group four (G4): highland mountain populations (over 1 800 m).

Additionally, 15 samples of *Funaria hygrometrica* were obtained from, Mediterranean zone (Greece, Italy and Turkey), Western Europe (United

Kingdom: England), North America (U.S.A.: North Carolina, California and Connecticut) and Asia (Russia, west Siberia) (Table 2.2).

Table 2.1 Locations sampled in the Sierra Nevada Mountains and surroundings. Locations are ordered by their altitude above sea level (a.s.l.) in meters (m), along with their description, positioning in geographic coordinates and number of collected samples.

Location altitude (m a.s.l.)	Description	
	Description and geographic coordinates	**Number of collected samples**
Loc_1 24 m	Road N-340 km 330, direction to Motril coming from Granada. The sample was found in a shady corner on the road side, near an artificial small water canal.	
	36° 43′ 57.4″N; 03° 31′ 00.9″W	1
Loc_2 58 m	Road N-340 km 334, direction to Motril coming from Granada. Samples were found in a palm nursery with a highly moist soil.	
	36° 43′ 47.4″N; 03° 30′ 30.8″W	9
Loc_3 58 m	Road N-340 km 370, direction to Motril coming from Granada. Samples were found on the road sides in a range of 30 m length.	
	36° 45′ 4.6″N; 03° 11′ 59.6″W	4
Loc_4 287 m	Road A-346 km 9, direction to Órgiva coming from Granada. Samples were collected on asphalt and soil slope on the edge of the road.	
	36° 51′ 56.5″N; 03° 28′ 59.6″W	22
Loc_5 382 m	Road A-346 km 7, direction to Órgiva coming from Granada. Samples were collected around the edge of the road.	
	36° 52′ 19.8″N; 03° 28′ 24.3″W	15
Loc_6 630 m	Lanjarón city. Samples were found on a soil slope on the edge of the road at the entrance of the city, the slope was nitrified.	
	36° 55′ 1.60″N; 03° 28′ 47.9″W	8
Loc_7 657 m	Out of Lanjarón city, road A-348, direction to Órgiva coming from Lanjarón. Samples were found on a soil slope on the edge of the road out of the city, the slope was nitrified and moist.	
	36° 55′ 1.40″N; 03° 28′ 17.0″W	14
Loc_8 755 m	Cenés de la Vega surroundings. Samples were found along the road and cliff edge in front of a restaurant.	
	37° 09′ 17.4″N; 03° 32′ 46.2″W	11
Loc_9 757 m	Road GR-420 km 1, direction to Pinos Genil coming from Cenés de la Vega. Samples were collected from the edge of the road in the proximity of a restaurant and farms.	
	37° 09′ 41.9″N; 03° 31′ 27.2″W	12

Continue Table 2.1

Location altitude (m a.s.l.)	Description	
	Description and geographic coordinates	**Number of collected samples**
Loc_10 1 295 m	Road A-395 km 15, up to Sierra Nevada crests. Samples were collected from the edge of the road, on soil slopes in the vicinity of La Higuera restaurant.	
	37° 08′ 34.1″N; 03° 29′ 14.5″W	12
Loc_11 1 328 m	Road A-395 km 16, up to Sierra Nevada crests. Samples were collected from the edge of the road, on soil slopes in the vicinity of Los Puentes restaurant.	
	37° 08′ 22.7″N; 03° 29′ 03.5″W	16
Loc_12 1 650 m	Road A-395, barranco de Las Víboras, up to Sierra Nevada crests. Samples were collected from the edge of the road, on soil slopes in the vicinity of Los Jamones restaurant.	
	37° 07′ 58.0″N; 03° 26′ 20.6″W	11
Loc_13 1 667 m	Road A-395, center for nature interpretation "El Dornajo". Samples were collected on soils in the proximity of the center.	
	37° 07′ 57.9″N; 03° 26′ 06.2″W	8
Loc_14 2 180 m	Ski resort Pradollano. Samples were collected on soils at the base of a chairlift machine and around sites.	
	37°05′ 35.7″N; 03° 23′ 54.5″W	2
Loc_15 2 220 m	Road A-395, military mountain hostel. Samples were collected from soil slopes around the hostel.	
	37° 06′ 47.9″N; 03° 25′ 10″W	2
Loc_16 2 622 m	Lake of Las Yeguas. Soil on rocks around the lake.	
	37° 04′ 22.3″N; 03° 23′ 28.2″W	3
Loc_17 2 700 m	Ski resort Borreguiles. The sample was collected on soil in the surroundings of a chairlift machine.	
	37° 04′ 16.4″N; 03° 23′ 14.6″W	1

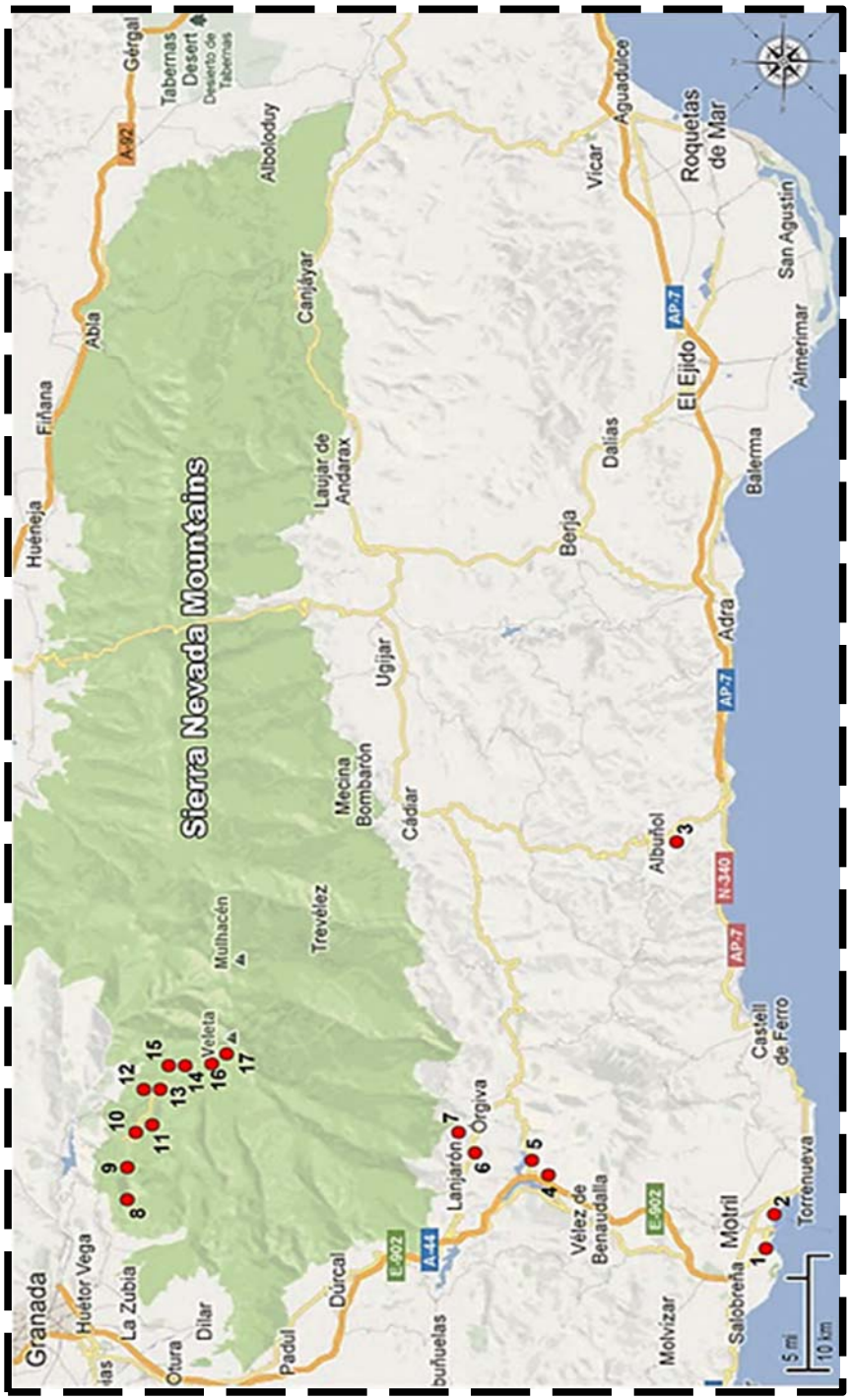

Figure 2.1 Sampling locations in Sierra Nevada Mountains of Spain and surroundings. The map was generated by Google earth™.

Table 2.2 *Funaria hygrometrica* samples from outside Sierra Nevada Mountains (Spain).

Country, state or region	Code	Herbarium (*)	Collection	Altitude (m a.s.l.)	Locality; habitat; year of collection
Greece, Lesvos	Greece 1	Herb. T.L. Blockeel	*Blockeel 38/141*	2	Eastern Aegean Island Lesvos, on the road to Alchaderi, 1.5 km south to the junction with Mitilini road; on gravelly ground on coastal mud flats; 1995
Greece, Lesvos	Greece 2	Herb. T.L. Blockeel	*Blockeel 38/115*	750-850	Lesvos, above and to south-west of Agiassos, higher slopes of Mt. Olympus; on soil on steep bank by forest road; 2009
Italy, Sicily	Sicily	MUB 42006	*Privitera s.n.*	1 320	Contrada Tassita, Parco dei Nebrodi; with great edafic and atmosferic humidity; 2009
Italy, Trentino Alto Adige	Italy	MUB 42007	*Tacchi s.n.*	1 400	Val di Rabbi; on soil; 2003
Russia, West Siberia	West Siberia	MUB 27215	*Czernyadjeva & Kuzmina s.n.*	-	Middle Ob River, vicinity of Surgut Town, lower Pochechujka River; flood plain of Ob River, on bare peat; 2000
Turkey, Denizli	Turkey 1	AYDN 2281	*Kirmaci 2997*	-	North of Honaz mountain, Yala position; *Pinus nigra* forest, limestone rock; 2005
Turkey, Aydin	Turkey 2	AYDN 1338	*Kirmaci 1165*	1 500	Aydin mountain, Pasha highland; half shade, moist soil rocks; 2003
United Kingdom, England	England	Herb. T.L. Blockeel	*Blockeel 40/558*	145	Derbyshire, 2 km north-east of Brailsford; on soil at egde of arable field; 2011
U.S.A., North Carolina	USA, NC 1	CONN	*Goffinet 9344*	115	Durham, Duke University Campus; edge of flower bed; 2006
U.S.A., North Carolina	USA, NC 2	CONN	*Goffinet 4738*	133	Chapel Hill, Lake Ellen; - ; -
U.S.A., California	USA, CA	CONN	*Villarreal 1284*	70	Merced County, San Luis reservoir; - ;-
U.S.A., Connecticut	USA, CN 1	CONN	*Goffinet 9278*	150	Town of Mansfield, at 219 Mount Hope Road; - ; -
U.S.A., Connecticut	USA, CN 2	CONN	*Budke 142*	67	Willimantic; - ; -
U.S.A., Connecticut	USA, CN 3	CONN	*Budke 143*	67	Willimantic; - ; -
U.S.A., Connecticut	USA, CN 4	CONN	*Budke 144*	67	Willimantic; - ; -

(*) AYDN: Herbarium of the Adnan Menderes University (Turkey); CONN: Herbarium of the University of Connecticut (U.S.A.); MUB: Herbarium of the University of Murcia (Spain).

PART 2.2

METHODS: LABORATORY PROTOCOLS

2.2.1 IDENTIFICATION

All plant samples were identified and confirmed using a Nikon® stereomicroscope and an Olympus microscope according to the morphological characters described in the literature (Smith 2004; Miller & Miller 2007).

2.2.2 CULTIVATION

The *in vitro* cultivation of *Funaria hygrometrica* was initiated from the mature capsules of the samples except for locations 14, 15, 16 and 17 as samples of these locations did not have viable spores. The main purpose of sample cultivation is to be sure that the genetic diversity corresponds to one sporophyte and the contamination by soil and natural microorganisms is eliminated. Plants with mature capsules, characterized by yellow-red colors were preferentially used to increase the cultivation success rate of the sample.

Plastic pre-sterile (disposable) Petri dishes were prepared with two different types of cultivation media. The first was a solid nutrient media Murashige and Skoog “MS” (Murashige & Skoog 1962), and the second was solid Knop media “KM” (Knop 1884).

A) MS hormone/sucrose-free media was prepared using the following protocol:
   1- 4.4 g of MS basal medium (SIGMA®, M5519-50L) were added to 800 ml of ddH_2O.
   2- Mix and adjust pH to 5.7 with 1 M KOH.
   3- Add 10 g of agar (Panreac, 141792).
   4- Add water to complete volume to 1 l.
   5- Autoclave for 20 to 25 min, and it was left at room temperature to be cooled to ~ 65 ^oC prior to be poured into Petri dishes (ca. 20 ml/dish).

B) Knop cultivation media preparation:

1- First, the following stock solutions were prepared and autoclaved:
   a. 25 g/l KH_2PO_4
   b. 25 g/l KCl
   c. 50 g/l $MgSO_4$ x 7 H_2O
   d. 100 g/l Ca $(NO_3)_2$
2- For 1 l Knop medium 10 ml of each stock solution were taken.
3- 12.5 mg $FeSO_4$ x 7 H_2O were added, and then pH adjusted to 5.8 with 1 M KOH.
4- 10 g of agar (Panreac, 141792) were added and the volume adjusted to 1 l.
6- The medium was sterilized by autoclaving 20 to 25 min, and it was left at room temperature to be cooled to ~ 65 ^oC prior to be poured into petri dishes (ca. 20 ml/dish).

Petri dishes containing solidified medium were stored in a sealed bag up to 4 weeks at room temperature, and used for cultivation after 3 days of storage as a contamination checkup test.

Sample's capsules and green gametophytes (samples of locations 14, 15, 16 and 17) were surface-sterilized for 5 min using 0.5 % sodium dichloroisocyanuric acid ($C_3Cl_2N_3NaO_3$). Capsules were then opened aseptically. Spores were transferred to petri dishes under sterile conditions.

All cultures were kept in incubator (BINDER KB 720, Germany) at 22±3 ºC, and 16/8 h of light to darkness, supplied by cool-white fluorescent tubes at a photon fluency rate of 33.5 to 55 µmol/sm^2 (Sabovljević *et al.* 2003).

The successfully uncontaminated germinated spores were sub-cultured at 15 days intervals, while contaminated plates were substituted and re-cultivated. Successfully cultivated protonemata were used for DNA extraction.

2.2.3 DNA EXTRACTION

Cultivated axenic protonemata and sterilized green apical portions of gametophores -in the cases where no viable spores were available- were used for DNA extractions. The DNA extractions were carried out using SIGMA®-ALDRICH GenElute™ Plant Genomic DNA Miniprep Kit (Cat# G2N350).

About 100 mg of plant material were frozen by soaking in liquid nitrogen, and then disrupted by grinding using mixer mill MM400 (RETSCH, Germany). After the preparation of the kit solutions, the kit protocol was followed without any modifications. Finally, the elute solution contained pure genomic DNA and was kept in 2 to 4 ^{o}C for immediate use, while kept in -20 ^{o}C for long-term storage. Multiple freezing and thawing circles were always avoided, as it causes DNA breaks.

DNA quality was checked using 1 % (w/v) agarose gel electrophoresis, visualized by pre-added RedSafe® dye under UV light and quantified using a spectrophotometer device.

2.2.3.1 Agarose gel electrophoresis protocol

This method was used to inspect the DNA degradation level for quality assurance. The protocol followed Tietz (1998):

1- Preparation of solutions:

a. 5 x TBE buffer (1 l)

Material	Concentration	To be added
Boric acid (MERCK, Germany)	0.45 M	27.50 g
Tris base (Invitrogen 15504-020)	0.45 M	54.00 g
EDTA-Na_2 (SIGMA E5134-500G)	12.50 mM	4.68 g
ddH_2O	-	up to 1 l

- Then the pH was adjusted to 8.3 using 1 M HCl

b. 10 x glycerol dye mix (10 ml)

Material	Concentration	To be added
Glycerol (MERCK, Germany)	50.00 %	5 ml
Bromophenol blue (SIGMA B8026)	0.05 %	5 mg
Xylene cyanol (SIGMA X4126)	0.05 %	5 mg
5 x TBE	1 x	2 ml
ddH_2O	-	up to 10 ml

2- Gel preparation:
   a. 1 g agarose powder was added to 100 ml of 1 x TBE buffer solution.
   b. The solution was heated in a microwave oven for 3 min.
   c. The solution was swirled gently to release trapped air and re-suspend any agarose particles caught on the side of the flask.
   d. The solution was heated again in a microwave oven for additional 1 min.
   e. 5 µl RedSafe® dye were added to the 100 ml solution and mixed gently to disperse the dye.
   f. The agarose solution was cooled down to 55 ^oC prior to be casted in the tray.
3- Gel casting:
   a. Warm agarose solution was poured into the gel tray with the comb already inserted.
   b. After the agarose had set for 30 min, gel tray was set in the buffer tank.
4- Electrophoretic device preparation:
   a. The device tank HU10 (Consort NV, UK) was filled with 1 x TBE buffer.
   b. The combs were gently removed by wiggling back and forth and then lifting up.
5- DNA sample preparation:
   a. DNA samples were prepared in microfuge tubes as 1 µl 10 x glycerol dye into a 10 µl DNA for each sample, and mixed gently.

b. Under the buffer, 5 µl were loaded to the gel, and ended with a positive sample and/or 2.5 µl of GeneRuler™ 100 bp DNA ladder (Fermentas, SM1143).

6- Electrophoresis:
   a. Based on 5 V for each cm between cathode and anode electrodes, the device was adjusted and run at 100 V. Ampere reading was used to determine the buffer stability and freshness.
   b. The electrophoresis ran for approximately 30 min, and stopped when the loading dye reached a pre-established position.

7- Gel visualization and documentation:
   a. The gel was illuminated with short wavelength UV light (254 nm) generated by mineral-light model R-52.
   b. The gel was digitalized using E-BOX-VX2/20M (VXBOX, France) gel documentation system.

2.2.3.2 Spectrophotometer test

This test was performed to quantify the DNA obtained from the DNA extraction step, which is important for DNA concentration sensitive techniques like AFLP-PCR technique. 100 µl of a 1:10 diluted DNA elution solution were prepared and measured for each sample using an Eppendorf® spectrophotometer x 100 device.

2.2.4 DNA SEQUENCING

2.2.4.1 PCR optimization

Initial PCR tests for all regions were established using 5 randomly selected samples from as far as possible separated locations to confirm DNA/primer amplification, purification and sequencing steps. When successful, sequences were aligned with a reference sequence to detect the overall variation ratio (variation test) by dividing the total number of variations in nucleotide sequence by the total length of the region. Successful reasonably variable candidate regions were used for sequencing all samples.

nrDNA primers. Initially the ITS region and the *FLORICAULA/LEAFY* homolog 1 gene (*PpLFY*1) were used for their evaluation. For the ITS region the primers of Douzery *et al.* (1999) were used with modifications introduced in order to match the published sequence of *Funaria hygrometrica* (GenBank: X74114). The ITS region was amplified completely as one region and then sequenced separately as ITS1 and ITS2. For the *FLORICAULA/LEAFY* region, the primers of Tanahashi *et al.* (2005) were redesigned based on the *Physcomitrella patens* genome data.

cpDNA primers. For the chloroplast DNA amplification, initially the intron region *trn*V primers (Werner *et al.* 2009), the *trn*K_{UUU} region primers (with *mat*K gene included; Borsch & Quandt 2009), *pet*N-*pet*D (Liu *et al.* 2012a) and the RNA polymerase Rpb1 (*rpo*C1) gene primers (Royal Botanic Gardens of Kew 2012) were tested. In the case of the *rps*3-*rpl*16 region (which includes the *rpl*16-*rps*3 spacer, the *rpl*16 exon I and the *rpl*16 intron), a completely new set of primers was designed. For this purpose the complete *Physcomitrella patens* chloroplast sequence (GenBank: NC005087) was used as the reference sequence. In a first step, four forward and three reverse primers were tested. Based on the obtained results, a new set of *Funaria*-specific primers was designed in order to improve the sequence quality.

mtDNA primers. The complete mitochondrion sequence of *Physcomitrella patens* obtained from GenBank (NC007945) was used as the reference sequence for comparing and testing published primers for *Funaria*

hygrometrica mtDNA amplification. Intergenic spacers *rpl*5-*rpl*16, *atp*1-*trn*W, *atp*7-*atp*6 and the *rpl*2 intron primers design were obtained from (Liu *et al.* 2012b), all primers were tested for confirmation and variation ratio. All primers were synthesized (Eurofins, Germany), and diluted for PCR reactions as described before. Primers are detailed in table 2.3, where the changes compared with the originally published primer sequences are written in bold.

Table 2.3 Region, length (bp), code, sequences and melting temperature (Tm) of the primers used in DNA sequencing. Changed nucleotides from reference sequences are indicated in bold.

Region	(bp) Primer*	5´-Sequence-3´	Tm ºC
ITS1	(28) 101-**F**-bryG	**C**CGA**T**T**GA**ATGGTCCGGTGA**G**GT**T**TTCG	65
	(20) 5.8s-**R**	GCTGCGTTCTTCATCGTTGC	60
ITS2	(20) 5.8s-**F**	GCAACGATGAAGAACGCAGC	60
	(31) 102-**R**-bryN	**GCTGGGCTC**T**TT**CCGGTTCGCTCGCCGTTAC	70
*PpLFY*1	(32) PpLFY-**F**2	CCCCGGGGTTTGAATGGTGGACGACTGCCCTG	72
	(32) PpLFY-**R**2	ATGTACCCGGGCTCATTCACCGTGCTTC	66
*trn*V	(16) trnV-**F**	ACACGTGCGCCAATGC	57
	(21) trnV-**R**	TYGAACCGTAGACATTCTCGG	55
*trn*K$_{UUU}$	(23) trnK-**F**	GGGTTGCTAACTCAATGGTAGAG	61
	(19) trnK-**R1**	GAACCCGGAACTHGTCGGAT	58
petN-petD	(19) petN-**F**2	CCATIAAAGCACCCCAAGC	57
	(19) petD-Fun-**R**	CCTTTCCGTCGTCCAGTAG	59
*rpoC*1	(23) 1-**F**	GTGGATACACTTCTTGATAATGG	54
	(23) 3-**R**	TGAGAAAACATAAGTAAACGGGC	58
*rps*3-*rpl*16	(22) rps3-**F**1	TAACAAAACCATACGGAGAACC	53
	(26) rps3-**F**2	GCTCGTGTAGAATGGGCTAGAGAAGG	60
	(34) rps3-**F**3	GGAGTATTAGGAATAAAAATTTGGATATTTCAAG	54
	(24) rpl16-**F**	TTGTTATGCTTAGTGTGCGACTCG	57
	(30) rpl16-**R**1	CGAAATTTTGTTCTTTTAGGGTTATAGTCG	55
	(25) rpl16-**R**2	CCTCGAGTAGATATTCCTTTCATCC	54
	(25) rpl16-**R**3	GCTCGTCGTCCTGCTTCTATTTGTC	59
*Funaria rps*3-*rpl*16	(22) rps3-**F**1-Fun	**CCAGCTCAAACAATTTATGGAG**	52
	(24) rpl16-**R**1-Fun	**CATTCTCCCTCTATGTTGTTTACG**	53
	(17) rpl16-**R**2-Fun	**GAATTACCTCGGGTAGC**	49
	(19) rpl16-**R**3-Fun	**AAGCAATAGAATTACCTCG**	47
*rpl*5-*rpl*16	(18) rpl16-**F**1	GGATGGTGTGAGTTTGTC	54
	(18) rpl5-**R**1	CGGAGTCTATTTGGAGTG	54
*atp*1-*trn*W	(15) atp1-**F**1	AGTCAACTGGCTACC	48
	(15) trnW-**R**1	ACAGGTTAAGGGTTC	45
*atp*7-*atp*6	(20) rps7-**F**1	TATRGCARCAAATCGTCAGA	53
	(20) atp6-**R**1	TATCTATGGGGGGTATTATG	53
*rpl*2	(18) rpl2-**F**1	AACACTTCGTCTATTGGC	51
	(18) rpl2-**R**1	GATACGCAATTTCCTACC	51

* **F** for forward primer, **R** for reverse primer.

Trials, optimization and all sample amplifications were performed under the following PCR conditions and steps:

1- A master mix was prepared and distributed depending on the PCR reactions number (sample/*locus*). Green GoTaq® Flexi DNA polymerase (Promega, M8295) was used in 50 µl reactions. In a sterile, nuclease-free micro-centrifuge tube, the following components were combined on ice:

Component	Stock concentration	Reaction concentration
GoTaq® Flexi buffer	5 x	1 x
$MgCl_2$	25 mM	4 mM
dNTPs	10 mM each	0.20 mM each
Forward primer	10 µM	1 µM
Reverse primer	10 µM	1 µM
Go Taq™	5 U/µl	1.25 U
DNA Stock	-	50 ng
a.ddH_2O	-	up to 50 µl total volume

2- The assay included blank (H_2O) in addition to negative and positive controls.

3- Micro-centrifuge tubes were placed in a Techne™ 96 thermocycler. The PCR program was designed considering what was described before.

   a. PCR general program:

Step	Temperature	Time	Cycles
Initial denaturation	95 ºC	2 min	1
Denaturation	95 ºC	30 sec	35
Annealing	55 to 65 ºC*	1 min	
Extension	72 ºC	30 sec + 1 sec /cycle	
Final extension	72 ºC	5 min	1
Soak (cooling)	04 ºC	indefinite	1

* Annealing temperatures were pre-tested and optimized for each region, but finally 60 ºC for ITS1 and ITS2 while 55 ºC for the *rps*3-*rpl*16 and the *rp*5-*rpl*16 regions were used.

4- PCR products were separated by 1.5 % (1.5 g / 100 ml) agarose gel electrophoresis and visualized with 1 x RedSafe® (5 µl / 100 ml) prepared as described before. 5 µl of PCR reactions contain the 1 x GreenGoTaq® Flexi buffer were loaded onto the gel directly after amplification, additionally one lane was loaded with 2.5 µl of

GeneRuler™ 100 bp DNA Ladder Plus (Fermentas, SM1153). Electrophoretic conditions and gel documentation were the same as described before (point 2.2.3.1).

2.2.4.2 PCR purification and elution

All sample PCR reactions, when successful, were prepared for the cleanup step (purification) in order to eliminate dimers, RNA, PCR residues and unamplified DNA by using GeneJET™ PCR purification kit (Fermentas, K0702). This It is an essential step before automated DNA-sequencing. The PCR product was bound to a silica membrane in the column, then washed twice and eluted in a new clean tube following the kit manufacturer's protocol.

The purification success was tested by 1.5 % (1.5 g / 100 ml) agarose gel electrophoresis and visualized with 1 x RedSafe® (5 µl / 100 ml) prepared as described before (point 2.2.3.1). 2 µl of purified DNA fragment mixed with 2 x Green GoTaq® Flexi buffer were loaded onto the gel, additionally one lane was loaded with 2.5 µl of GeneRuler™ 100 bp DNA Ladder Plus (Fermentas, SM1153). Electrophoretic conditions were the same as described before (point 2.2.3.1).

2.2.4.3 PCR sequencing

After successful purification, the purified DNA was prepared to be sent to the sequence service (Secugen, Spain or Macrogen, Holland) to proceed with the sequencing process.

To sequence all samples for each selected region, 96-well PCR-plates were prepared and both strands of each sample were sequenced. Each sample was prepared by adding 1 µl of primer (10 µM) to 10 µl of purified DNA product (approx. 50 µg).

2.2.4.4 Sequence alignment

For each region, the sequence chromatograms were compiled using BIOEDIT V3 (Hall 1999), which facilitates a working environment that allows viewing and manipulating sequences with simple point-and-click operations to assemble the sequences in the following steps:

1- Sequence service chromatogram files (.fas) of each-sample (i.e. one forward and one reverse) were imported into a new alignment file created in BIOEDIT.
2- The reverse sequence was first converted into reverse-complement sequence and aligned automatically with the forward sequence using the option ClustelW (Thompson *et al.* 1994) to detect and confirm variations. Whenever a variation between the two sequences was detected, the chromatogram signals (trace file) were used to decide which nucleotide type to assign based on the highest stable peak criteria. Finally one copy of the two sequences (5´ → 3´ direction) was used for between-samples multiple alignment.
3- The final copies of the samples were aligned manually together in a new alignment file, while gaps were inserted to preserve nucleotide homology and fragment length.
4- Intragenic boundaries for each region were determined using reference sequences mentioned before to delete extra nucleotides and the trimmed alignment was saved in “fasta” format in order to proceed with the analysis step.
5- All haplotype sequences were submitted to the GenBank database using the web built in “BankIt Submission Tool” provided by the National Center for Biotechnology Information (NCBI).

2.2.5 FINGERPRINTING TECHNIQUES

2.2.5.1 Microsatellites set up and optimization

Dr. Bernard Goffinet (University of Connecticut, U.S.A.) shared information of on 20 unpublished SSR *loci* and (forward "F") with (reverse "R") primers for each (Table 2.4).

Table 2.4 SSR *loci* information shared by Dr. Bernard Goffinet. SSR *loci*, primers sequence, melting temperature (TM), repeats, expected fragment size (bp) and blast results, are shown.

Loci	5´- Forward/Reverse - 3´	Tm ºC*	Repeats	bp	blast result
M1	GAGTCGCTAATACGGGTGC	55.55	(AT)^13	449	*Funaria* cp genome
	CCAAACTGCGCTACATCCC	56.4			
M2	TAGCCTTTGACTGGCTCCC	57	(AT)^12	269	*Funaria* mt genome
	TGGGCAGGATTTACCCATC	54.7			
M3	GAGGCGCCATTCAAGATCC	56.3	(GT)^12	434	*Physcomitrella patens*
	GTTGCTTGGGCTGAAGGAC	56.7			
M4	GTGCAAGCACTTCGATCCC	56.6	(GT)^13	389	*Physcomitrella patens*
	TGCATTCTGACACACAGTTCG	55.6			
M5	AGATGAAGCCATGAGGCCG	57.6	(AT)^11	351	Zebrafish
	GGCTTGCCATCCTTAATGGG	56.4			
M6	GTCAGCGAGCACTGAAAGC	56.7	(AT)^10	422	*Oryza sativa japonica*
	ATCCTAGTGTGGGAGACGC	56.3			
M7	TCTGGCGAGTGGCTTATCG	57.1	(AC)^12	305	*Physcomitrella patens*
	CAAATGGGATGCACTTGTTGAC	55			
M8	CGCATGCTGGAATGTGGAC	57	(AC)^11 (AT)^10	341	*Physcomitrella patens*
	GCTCTCCAACAAAGCCCAG	56.4			
M9	TCTGAGCCATTCCACGGAG	57.8	(GT)^11	377	Mouse, human
	GTGCCACATTCCGTGATCC	56.4			
M10	AGCCATAGGAAATGAAACTTCTCG	55.1	(AT)^13	443	Zebrafish
	GGCTCTCATGGCCACTTTG	56.5			
M11	TCCCACGAACCCATGTTCC	57.5	(AC)^10	445	*Pinus*
	TCTGAGCCATTCCACGGAG	56.9			
M12	CCCATCGCCCATAACTGAAG	55.7	(AC)^11	334	Zebrafish
	AAATGCCAAGGTGTGGGAC	55.9			
M13	TTGCTTTGCTAGGGCATGG	56	(GT)^10	194	Mouse, human
	GGAATTTGGAGACCCAGCG	56.2			
M14	TCTACCACCACCTGTTCGC	57	(AG)^10	252	*Physcomitrella patens*
	TGGACTGCAGATGGACAGC	57.5			
M15	TGTGCTTCACAAACCGTGC	56.6	(AT)^11	222	Zebrafish
	TAGGCAGAACTTGGAGCGG	57.1			
M16	TTCACGCAATCGTGGACCC	58.1	(CT)^10	309	*Physcomitrella patens*
	GACCCGCGTCGAGAACC	58.7			
M17	CTGTGCACTCAATTTCATACGTG	54.5	(AT)^10	213	*Physcomitrella patens*
	ACGTCCATGTTTGATGAGACAAC	55.5			
M18	TCTCAAACGCAACTTCTGC	53.4	(AG)^10	416	Zebrafish
	CACAACCGTCCTTCGCAC	56.8			
M19	CACGACTCGGGATAGGGTC	56.9	(CT)^11	151	*Physcomitrella patens*
	CGCGCGGCTTTGACATC	57.9			
M20	GTGTTCCTTCACCTGGTCTTG	55.7	(CT)^10	188	Mouse, human
	GTTGTGAGACGAGGGCTCC	57.9			

* Tm was determined after primer synthesis

All primers were synthesized (Eurofins, MWG, Operon, Germany) and re-suspended to 100 µM with autoclaved double distilled water (a.ddH_2O). Working solutions were diluted to 10 µM for PCR use. Re-suspended stocks were kept at -20 ºC while the diluted tubes were kept at 4 ºC for immediate use.

PCRs were performed to test the amplification ability of the 20 SSR *loci* using only 15 samples that were randomly selected. PCR reactions were performed under the following conditions and steps:

1- A master mix was prepared with all the necessary components for the PCR reaction with the exception of the DNA. Green GoTaq® Flexi DNA polymerase (Promega, M8295) was used. The final volume of the sample reactions was 25 µl. The reaction for each *locus* was set up individually (no multiplexing) as follows:

Component	Stock concentration	Reaction concentration
GoTaq® Flexi buffer	5 x	1 x
MgCl2	25 mM	4 mM
dNTPs	10 mM each	0.20 mM each
Forward primer	10 µM	1 µM
Reverse primer	10 µM	1 µM
Go Taq™	5 U/µl	2 U
DNA stock	variable	50 ng
a.ddH_2O	-	up to 25 µl total volume

2- The assay included a blank (H_2O) to evaluate the rate of evaporation, along with negative (mix without DNA) and positive (mix with previously amplified DNA) controls.
3- Micro-centrifuge tubes were placed in a Techne™ 96 thermocycler, and the PCR program was designed considering the following:
    a. The annealing temperature was set to intermediate Tm value for each primer pair according to the Tm given by the manufacturer, in case if the Tm of the primer pair is approximately equal ± 1 ºC; otherwise the annealing temperature was set to the lower Tm value.

b. As recommended by Promega©, the thermal cycling protocol has an initial denaturation step where samples are heated at 95 °C for 2 min to ensure that the target DNA is completely denatured. Initial denaturation of longer than 2 min at 95°C is unnecessary and may reduce the yield.
c. As recommended by Promega©, the extension time was set to be at least 1 min/kb target length.
d. Extension started with 30 sec and increased + 1 sec per cycle.
e. PCR microsatellite general program:

Step	Temperature	Time	Cycles
Initial denaturation	95 ºC	2 min	1
Denaturation	95 ºC	30 sec	30
Annealing	53 to 59 ºC*	1 min	
Extension	72 ºC	30 sec + 1/cycle	
Final extension	72 ºC	5 min	1
Soak (cooling)	04 ºC	indefinite	1

* Annealing temperature optimized for each primer set based on the Tm.

4- PCR products were separated by 1.5 % (1.5 g / 100 ml) agarose gel electrophoresis and visualized with 1 x RedSafe® (5 µl / 100 ml) prepared as described before (point 2.2.3.1). 5 µl of PCR reactions containing the 5 x GreenGoTaq® Flexi buffer were loaded onto the gel directly after amplification, additionally one lane was loaded with 2.5 µl of GeneRuler™ 100 bp DNA Ladder Plus (Fermentas, SM1153). Electrophoretic conditions and gel documentation were the same as described before (point 2.2.3.1).

2.2.5.2 Selected SSR-PCR

Five SSR *loci* were selected to amplify all samples, based on their polymorphic level and consistent amplification, PCR was performed using the program described before (point 2.2.5.1). Samples were visualized using polyacrylamide gel electrophoresis (PAGE), and submerged in RedSafe® dye solution, and finally documented as described before (point 2.2.3.1).

Polyacrylamide gel electrophoresis (PAGE) is more sensitive in separating short DNA fragments than agarose gel electrophoresis. The protocol from Sambrook *et al.* (1989) was applied with some minor changes:

1- Solutions preparation:

a. 30 % polyacrylamide stock solution:

Material*	To be added
Acrylamide (Scharlau)	29 g
Bis-acrylamide (Scharlau)	1 g
ddH_2O	up to 100 ml

* The acrylamide and bis-acrylamide were dissolved in distilled water for 1 h, filtered through a 0.4 µm filter to remove un-dissolved particles; finally stored in a dark bottle at 4 ^oC.

b. 10 % ammonium persulfate

Material	To be added
Ammonium persulfate (Panreac, 131138)	100 mg
ddH_2O	up to 1 ml

c. 1 x RedSafe® dye solution

Material	To be added
20 000 x RedSafe® dye (iNtRON, 21141)	50 µl
ddH_2O	up to 1 l

2- Gel preparation:

a. The gel casting glass and tray of the mini-gel RELSVS10WD (REAL, UK) device were washed, cleaned and fixed.

b. A 8 % polyacrylamide gel solution was prepared according to the following instructions:

Material	To be added
30 % acrylamide (29:1)	3.2 ml
5 x TBE	2.4 ml
10 % APS	200 µl
N,N,N',N' - tetramethylenediamine TEMED (SIGMA, T7024)	10 µl
ddH_2O	up to 12 ml

3- Gel casting:
    a. Immediately, the gel mixture was poured slowly into the previously prepared gel glass, until the liquid level reaches the top of the upper glass plate.
    b. The comb inserted between the two glass plates and the built in clampers where tightened well.
    c. The gel was left 1 to 2 h to solidify.
4- Electrophoretic device preparation:
    a. The device tank was filled with 1 x TBE buffer up to the max fill sign.
    b. The combs were gently removed, followed by washing wells with ddH_2O.
    c. Gel was placed gently into the tank and the trapped air bubbles below the gel were removed.
5- DNA sample preparation:
    a. DNA samples were prepared in microfuge tubes by adding 1 µl 10 x glycerol dye to a 9 µl DNA for each sample, and mixed gently.
    b. 2.5 µl of the dye-DNA-mix were loaded on the gel under the buffer, the GeneRuler^{TM} 100 bp DNA Ladder Plus (Fermentas, SM1153) was added as a size marker.
6- Electrophoresis:
    a. Pre-electrophoresed before use were done on 4 V/cm for 30 min, in order to eliminate charged impurities (assumed to be ammonium per-sulfate derivatives) from the gel matrix. Charged impurities will form complexes with the DNA fragments, changing their electrophoretic mobility in a non-reproducible manner.
    b. The device was adjusted and ran on low volts (4 V/cm) to prevent heat denaturation effects on the DNA bands. Ampere reading was used to determine the buffer stability and freshness.
    c. The electrophoresis ran for approximately 3 h, and was stopped when the xylene cyanol dye reached the end of the gel.

7- Gel visualization and documentation:
    a. The gel glass was submerged in the 1 x RedSafe® dye solution for 1 h; then soaked in ddH_2O for 30 min before visualization by UV light.
    b. The gel was carefully placed into the gel documentation system UV room equipped with UV light of short wavelength (254 nm) generated by mineral-light model R-52.
    c. The gel was digitalized and saved by E-BOX-VX2/20M (VXBOX, France) gel documentation system.

2.2.5.3 Fluorescently labeled SSR-PCR

A total of five forward primers out of the 20 primer pairs were chosen according to the optimization step to be re-synthesized and labeled fluorescently with "5´-HEX, 5´-FAM and 5´-CY3" (Habera *et al.* 2004). The fluorescently labeled forward primers were re-suspended to 100 µM with (a.ddH_2O), and working solutions were prepared with a final concentration of 10 µM for PCR use (Table 2.5). Re-suspended stocks were kept at -20 ^oC while the diluted tubes were kept at 4 ^oC for immediate use.

Table 2.5 Fluorescently labeled SSR *loci*. Fluorescence label type, labeled forward primers and their melting temperature (Tm-1), reverse primers and their melting temperature (Tm-2) and the PCR annealing temperature, are shown.

Label type	Forward primer	Tm-1	Reverse primer	Tm-2	PCR Tm
FAM	F2	58.8	R2	54.7	54.7
HEX	F3	58.8	R3	56.7	56.7
FAM	F12	59.4	R12	55.9	55.9
HEX	F13	56.7	R13	56.2	56.5
CY3	F14	58.8	R14	57.5	57.5

The SSR-PCRs were amplified, tested using agarose gel electrophoresis, visualized by RedSafe® and documented as described before (point 2.2.3.1).

Amplified products were visualized by Secugen, S.L. sequencing service (Spain), where an ABI3730 DNA analyzer (Applied Biosystems, U.S.A.) was used with a size standard (RoxGeneScan500) to analyze SSR amplified fragments.

The process was optimized by dilution calibration plate that was essential for the optimization of the fragment fluorescent emission with the standard size emissions. Three different amplified *loci* of eight randomly selected samples were mixed together. Each of them was loaded with different dilution ratios (1:1, 1:2, 1:5, 1:10, 1:15, 1:20) using a.ddH_2O. The optimal dilution ratio was then used for the final sample preparation of the complete set.

2.2.5.4 AFLP set up and optimization

For AFLP, the original protocol of Vos *et al.* (1995) was used, while modifications suggested by various authors were followed (James *et al.* 2003; Fernandez *et al.* 2006; Kamisugi *et al.* 2008; Huang & He 2010; Sonibare *et al.* 2010; Song *et al.* 2011; Mikulášková *et al.* 2012). Adaptors and primer sets were all synthesized by Eurofins (Germany) according to the original protocol as shown in Table 2.6.

Table 2.6 Sequence 5´- 3´of primers and adaptors used to establish the AFLP technique according to the original protocol of Vos *et al.* (1995).

Primer/Adaptor	5´ - sequence - 3´
EcoRI - A1	CTCGTAGACTGCGTACC
EcoRI - A2	AATTGGTACGCAGTC
Mse I - A1	GACGATGAGTCCTGAG
Mse I - A2	TACTCAGGACTCAT
Eco + A	GACTGCGTACCAATTC**A**
Mse + C	GATGAGTCCTGAGTAA**C**
Eco + ACA	*FAM*-GACTGCGTACCAATTC**ACA**
Eco + AGG	*HEX*-GACTGCGTACCAATTC**AGG**
Eco + ATA	*CY3*-GACTGCGTACCAATTC**ATA**
Mse + CAA	GATGAGTCCTGAGTAA**CAA**
Mse + CTC	GATGAGTCCTGAGTAA**CTC**
Mse + CAT	GATGAGTCCTGAGTAA**CAT**
Mse + CTA	GATGAGTCCTGAGTAA**CTA**

The established protocol according to the original three steps was the following:

1- Preparation of adaptors and primers:
   a. Adaptors preparation. The adaptors were re-suspended to 100 µM with a.ddH_2O. A working solution of 10 µM was prepared for PCR use and kept at 4 ^oC for immediate use.
      i. For EcoRI adaptor pair (final concentration of 5 µM):

Material	Concentration	To be added
EcoRI forward adaptor	100 µM	25 µl
EcoRI reverse adaptor	100 µM	25 µl
a.ddH_2O	-	450 µl

      ii. For Mse I adaptor pair (final concentration of 50 µM):

Material	Concentration	To be added
Mse I forward adaptor	100 µM	250 µl
Mse I reverse adaptor	100 µM	250 µl

      iii. After mixing the adaptors, they were heated at 95 ^oC for 5 min to denature and then left to cool slowly at room temperature to re-nature completely for at least 1 h.
   b. Primers preparation. The primers were re-suspended to 100 µM with a.ddH2O, and working solutions were prepared of 10 µM for PCR use and kept at 4 ^oC for immediate use.
2- Restriction “R” and ligation “L” of adaptors:
   a. Preparation of RL mix. 10 µl of DNA were added to a mix containing the restriction enzymes Mse I (BioLabs, R0525S), EcoRI (Fermentas, ER0271), both Mse I and EcoRI adaptors, ATP (Fermentas, R0441), T4 ligase (Fermentas, EL0015) and Tango buffer (Fermentas, BY5) that allows the activity of all the added enzymes. Finally a.ddH_2O was added up to 20 µl total volume:

Component	Stock concentration	Reaction concentration
MseI enzyme	10 U/µl	5 U
EcoRI enzyme	10 U/µl	5 U
MseI adaptor	50 mM	5 mM
EcoRI adaptor	5 mM	0.5 mM
ATP	100 mM	0.2 mM
T4	5 U/µl	1 U
Tango buffer	10 x	2 µl
DNA solution	variable	10 µl
a.ddH_2O	-	up to 20 µl total volume

b. The RL reactions were incubated at 37 °C for 4 h, and then placed on ice.

c. The RL reactions were tested by 1 % (1 g / 100 ml) agarose gel electrophoresis and visualized with 1 x RedSafe® (5 µl / 100 ml) prepared as described before (point 2.2.3.1). 5 µl of RL reactions containing the 5 x GreenGoTaq® Flexi buffer were loaded onto the gel, additionally one lane was loaded with 2.5 µl of GeneRuler™ 100 bp DNA Ladder Plus (Fermentas, SM1153). Electrophoretic conditions were the same as described before (point 2.2.3.1).

3- Pre-selective PCR. PCR reactions were performed using the adaptor specific primer with one base extension. PCR reactions were performed under the conditions described before with the following modifications:

a. A master mix was prepared and distributed depending on the number of PCR reactions. Green GoTaq® Flexi DNA polymerase (Promega, M8295) was applied into 25 µl mix/reaction. In sterile, nuclease-free micro-centrifuge tubes, the following components were combined on ice:

Component	Stock concentration	Reaction concentration
GoTaq® Flexi buffer	5 x	1 x
$MgCl_2$	25 mM	4 mM
dNTPs	10 mM	0.2 mM
Eco+A primer	10 µM	1 µM
Mse+C primer	10 µM	1 µM
Go Taq™	5 U/µl	1.25 U
RL mix	-	0.2 µl
a.ddH_2O	-	up to 25 µl total volume

b. 0.2 µl RL mix for every 25 µl pre-selective PCR reaction.

c. Micro-centrifuge tubes were placed in a FlexCycler (Analytic Jena) 96 thermocycler. Negative and positive controls were included.

d. Pre-selective PCR program:

Step	Temperature	Time	Cycles
Pre-denaturation	95 ºC	45 sec	1
Pre-annealing	65 ºC	30 sec	
Pre-extension	72 ºC	1 min	
Denaturation	95 ºC	30 sec	20
Annealing	56 ºC	30 sec + 1 sec/cycle	
Extension	72 ºC	30 sec	
Final extension	72 ºC	10 min	1
Soak (cooling)	04 ºC	indefinite	1

e. PCR reactions were tested by 1 % (1 g / 100 ml) agarose gel electrophoresis and visualized with 1 x RedSafe® (5 µl / 100 ml) prepared as described before (point 2.2.3.1). 5 µl of PCR reactions containing the 5 x GreenGoTaq® Flexi buffer were loaded onto the gel directly after amplification, additionally one lane was loaded with 2.5 µl of GeneRuler™ 100 bp DNA Ladder Plus (Fermentas, SM1153). Electrophoretic conditions were the same as described before (point 2.2.3.1).

f. Pre-selective PCR reactions were diluted 1:1 with a.ddH_2O and kept at 4 ºC for immediate use, or at -20 ºC for next day amplifications.

4- Selective PCR. PCR reactions were performed using the adaptor specific primers (Eco + 3 and Mse + 3). A total of 12 different combinations were amplified in the following order:

Combination code	Eco + 3 primers	Mse + 3 primers
C1	Eco + ACA	Mse + CAA
C2		Mse + CTC
C3		Mse + CAT
C4		Mse + CTA
C5	Eco + AGG	Mse + CAA
C6		Mse + CTC
C7		Mse + CAT
C8		Mse + CTA
C9	Eco + ATA	Mse + CAA
C10		Mse + CTC
C11		Mse + CAT
C12		Mse + CTA

PCR reactions were performed as follows:

a. A master mix was prepared and distributed depending on the number of PCR reactions. Green GoTaq® Flexi DNA polymerase (Promega, M8295) was used, with a final volume of 25 µl mix/reaction. In sterile, nuclease-free micro-centrifuge tubes the master mix was distributed. The following components were combined on ice:

Component	Stock concentration	Reaction concentration
GoTaq® Flexi buffer	5 x	1 x
MgCl2	25 mM	4 mM
dNTPs	10 mM	0.2 mM
Fluo-Eco + 3 primer	10 µM	1 µM
Mse + 3 primer	10 µM	1 µM
Go Taq™	5 U/µl	1.25 U
Pre-selective mix	-	0.2 µl
a.ddH_2O	-	up to 25 µl total volume

b. The assay included blank (H_2O), in addition to negative and positive controls.

c. Micro-centrifuge tubes were placed in a FlexCycler (Analytik Jena) 96 thermocycler.

d. Selective touch-down (TD) PCR program was the following:

Step	Temperature	Time	Cycles
Pre-denaturation	95 ºC	45 sec	1
Pre-annealing	65 ºC	30 sec	
Pre-extension	72 ºC	1 min	
TD denaturation	95 ºC	30 sec	12
TD annealing	64.3 ºC - 0.7 ºC/cycle	30 sec	
TD extension	72 ºC	30 sec + 1 sec/cycle	
Denaturation	95 ºC	30 sec	20
Annealing	56 ºC	30 sec	
Extension	72 ºC	30 sec + 1 sec/cycle	
Final extension	72 ºC	10 min	1
Soak (cooling)	04 ºC	indefinite	1

e. PCR reactions were tested by 1 % (1 g / 100 ml) agarose gel electrophoresis and visualized with 1 x RedSafe® (5 µl / 100 ml) prepared as described before (point 2.2.3.1). 5 µl of PCR reactions containing the 5 x GreenGoTaq® Flexi buffer were loaded onto the gel directly after amplification, additionally one lane was loaded with 2.5 µl of GeneRuler™ 100 bp DNA Ladder

Plus (Fermentas, SM1153). Electrophoretic conditions and gel documentation were the same as described before (point 2.2.3.1).

f. Selective PCR reactions were prepared to be sent for multiplex fragment analysis service (Secugen, Spain) in the following steps:

   i. Three differently labeled PCR products of the same sample were mixed together (triplex analysis), as follows:

Selective PCR product	**To be added**
FAM labeled	2 µl
HEX labeled	2 µl
CY3 labeled	2 µl
a.ddH_2O	4 µl
Total	10 µl

   ii. All samples were multiplexed and sorted by order in a sterile 96-well PCR-plate and sent to the fragment analysis service.

2.2.6 CLIMATIC DATA EXTRACTION

Climate layers were obtained from WorldClim database version 1.4 (WorldClim - Global Climate Data 2012). Methods used to generate the climatic layers, the units and formats of the data were set by Hijmans *et al.* (2005). WorldClim creates global climatic layers through interpolation of average monthly climatic data from weather stations around the world. Current climate data are based on averages from the years 1960-2000.

For current climatic layers, WorldClim provides a set of 19 bioclimatic variables (Table 2.7), which are derived from the monthly temperature and rainfall values in order to generate biologically meaningful variables (Hijmans *et al.* 2005). The bioclimatic variables represent annual trends (e.g., mean annual temperature, annual precipitation) seasonality (e.g. annual range in temperature and precipitation) and extreme or limiting environmental factors (e.g. temperature of the coldest and warmest month, and precipitation of the wet and dry quarters). The numerical values of each bioclimatic variable represent the average over a year (Hijmans *et al.* 2005).

Climatic variables were considered from the same climate layers of WorldClim database (Table 2.8). They include monthly averages of maximum temperature, minimum temperature and precipitation. As *Funaria hygrometrica'*s life cycle is completed in approximately 4 months (Nakosteen & Hughes 1978), only the data from February to June were considered with the aim to include months when samples were supposed to complete one life cycle in the study area (1st collection on May 2010 and 2nd collection in July 2011).

Both kinds of variables were downloaded directly from the website as generic grids at 30 arc-second resolution (0.93 km x 0.93 km = 0.86 km^2 at the equator). Masks of the study areas (Sierra Nevada Mountains) were applied to the layers to minimize processing time and storage space.

For the bioclimatic and climatic data extraction, all of available WorldClim layers for temperature and precipitation were loaded to DIVA-GIS

v7.4.0.1 (Hijmans *et al.* 2001). The data processing to obtain the Euclidean distance matrix was the following:

1- Sampling data including name, location, coordinate and altitudes were formatted according to the program manual using Microsoft® EXCEL 2010.
2- Climatic database, geographic layers and sampling data were loaded to a new project named “Funaria Sierra Nevada”. The sampling locations coordinates’ column was used as plot-on-layer data.
3- On each plot point (location), the climatic data were extracted by click and copy, from all the layers, assigned as bioclimatic variables (Hijmans *et al.* 2001).
4- The bioclimatic and climatic data were pasted by location order in Microsoft® EXCEL 2010 as shown in tables 2.7 and 2.8. Afterwards, each variable was standardized using PCO3 (Anderson 2003) by subtracting each value from the sample mean and dividing the result by the sample standard deviation, known as z-score transformation.
5- Using PCO3 features, a principal component analysis (PCA) was performed in order to test if the correlations between the bioclimatic variables allow them to be treated as a single unit or needed to be separated into sets.
6- In the PCA plot, when more than one variable grouped together, they were defined as a set. Sets were numbered according to their size in a descending form (i.e. the set with the most number of variables were defined as “set1”). Unique variables were defined as singletons.
7- By using the option “Calculating geographic distance”, the difference in km between every two locations was formatted into a matrix form using Microsoft® EXCEL 2010.
8- Squared Euclidean distance matrices for bioclimatic variables sets and singletons, climatic variables, and altitudinal gradient were calculated from the standardized values between every two locations by using the following formula:

- The simple Euclidean distance between two points (p) and (q) is the length of the line segment connecting them (Deza & Deza 2009)

$$d\,(p,q) = d\,(q,p) = \sqrt{\sum_{i=1}^{n}(qi - pi)^2}$$

where n is the total number of locations

9- A Mantel test was performed between bioclimatic variables sets and singletons Euclidean matrices, in order to associate singletons to sets and to improve variables definitions.

10- When successful, step 8 was repeated for the newly defined set.

Table 2.7 19 Bioclimatic variables and their values for 17 sampling locations in Sierra Nevada Mountains (Spain). Bioclimatic variables are numbered and ordered as extracted from the database (Hijman *et al.* 2005). Location numbers and altitude (m a.s.l.) are shown.

	Location	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
	Altitude (m a.s.l.)	24	58	58	287	382	630	657	755	757	1 295	1 328	1 650	1 667	2 180	2 200	2 622	2 700
Bioclimatic variables	Annual mean temperature [1]	18.30	18.30	17.20	16.00	16.00	12.90	12.90	14.00	14.00	12.70	12.70	12.00	12.00	7.90	9.30	4.90	4.90
	Mean monthly temperature range [2]	9.00	9.00	9.40	10.60	10.60	11.80	11.80	12.00	12.00	12.00	12.00	12.10	12.10	12.30	12.20	12.40	12.40
	Isothermality [3]	38.40	38.40	38.00	38.60	38.60	38.50	38.50	38.20	38.20	38.10	38.10	38.10	38.10	38.00	38.10	37.90	37.90
	Temperature seasonality [4]	522	522	549	597	597	654	654	672	672	676	676	678	678	691	685	701	701
	Max temperature of warmest month [5]	30.80	30.80	30.70	31.50	31.50	30.90	30.90	32.50	32.50	31.40	31.40	30.80	30.80	27.40	28.60	24.80	24.80
	Min temperature of coldest month [6]	7.40	7.40	6.10	4.00	4.00	0.20	0.20	1.00	1.00	-0.20	-0.20	-0.90	-0.90	-4.90	-3.50	-7.90	-7.90
	Temperature annual range [7]	23.40	23.40	24.60	27.50	27.50	30.70	30.70	31.50	31.50	31.60	31.60	31.70	31.70	32.30	32.10	32.70	32.70
	Mean temperature of wettest quarter [8]	13.20	13.20	11.90	10.10	10.10	6.50	6.50	7.40	7.40	6.10	6.10	7.20	7.20	2.90	4.40	-0.20	-0.20
	Mean temperature of driest quarter [9]	24.80	24.80	24.10	23.70	23.70	21.40	21.40	22.80	22.80	21.60	21.60	20.90	20.90	17.00	18.30	14.20	14.20
	Mean temperature of warmest quarter [10]	25.00	25.00	24.40	23.70	23.70	21.40	21.40	22.80	22.80	21.60	21.60	20.90	20.90	17.10	18.40	14.30	14.30
	Mean temperature of coldest quarter [11]	12.20	12.20	10.90	9.20	9.20	5.50	5.50	6.40	6.40	5.20	5.20	4.40	4.40	0.40	1.80	-2.60	-2.60
	Annual precipitation [12]	312	312	310	395	395	567	567	521	521	584	584	624	624	861	775	1058	1058
	Precipitation of wettest month [13]	46	46	43	54	54	72	72	69	69	74	74	78	78	105	96	126	126
	Precipitation of driest month [14]	2	2	2	4	4	9	9	7	7	9	9	11	11	23	17	35	35
	Precipitation seasonality [15]	64.00	64.00	58.80	57.50	57.50	52.40	52.40	53.80	53.80	51.50	51.50	49.80	49.80	43.10	46.50	39.10	39.10
	Precipitation of wettest quarter [16]	135	135	124	157	157	207	207	190	190	207	207	220	220	301	274	365	365
	Precipitation of driest quarter [17]	11	11	13	19	19	37	37	30	30	39	39	46	46	87	68	125	125
	Precipitation of warmest quarter [18]	16	16	19	26	26	37	37	30	30	39	39	46	46	91	74	125	125
	Precipitation of coldest quarter [19]	127	127	119	150	150	204	204	188	188	206	206	217	217	281	261	332	332

Table 2.8 Climatic variables and their values of five months for 17 sampling locations in Sierra Nevada Mountains (Spain). Minimum temperature (Tmin), maximum temperature (Tmax), precipitation (P), location numbers, location altitudes (m a.s.l.) and average value of 5 months for each climatic variable *per* location, are shown.

	Location		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
	Altitude (m a.s.l.)		24	58	58	287	382	630	657	755	757	1 295	1 328	1 650	1 667	2 180	2 200	2 622	2 700
Climatic variables	Maximum temperature (Tmax)	February	16.9	16.9	15.7	14.6	14.6	11.2	11.2	13.4	12.3	10.9	10.9	10.9	10.1	8.5	6	4.6	1.8
		March	18.8	18.8	17.7	16.7	16.7	13.4	13.4	15.9	14.7	13.3	13.3	13.3	12.5	12.5	9.6	8.1	4.9
		April	20.7	20.7	19.7	18.9	18.9	16.1	16.1	18.5	17.4	16	16	16	15.2	15.2	12.5	11	7.9
		May	23.7	23.7	22.6	22.2	22.2	19.5	19.5	21.9	20.9	19.6	19.6	19.6	18.8	18.8	16.2	14.8	11.8
		June	27.8	27.8	27.2	27.6	27.6	25.8	25.8	28.7	27.6	26.2	26.2	26.2	25.4	25.4	22.7	21.3	18.4
		Average	21.58	21.58	20.58	20	20	17.2	17.2	19.68	18.58	17.2	17.2	17.2	16.4	16.08	13.4	11.96	8.96
	Minimum temperature (Tmin)	February	8.2	8.2	6.7	4.8	4.8	0.8	0.8	2.8	1.8	0.5	0.5	0.5	-0.3	-0.9	-3.5	-4.9	-7.9
		March	10.2	10.2	8.9	7.1	7.1	3	3	5.2	4.1	2.7	2.7	2.7	1.8	1.8	-1.1	-2.6	-5.9
		April	11.8	11.8	10.5	8.7	8.7	4.8	4.8	7.1	6	4.6	4.6	4.6	3.7	3.7	0.8	-0.7	-3.8
		May	14.5	14.5	13.3	11.5	11.5	7.6	7.6	9.8	8.7	7.4	7.4	7.4	6.6	6.6	3.7	2.2	-0.8
		June	18	18	16.9	15.3	15.3	11.7	11.7	14.1	13	11.6	11.6	11.6	10.8	10.8	7.9	6.5	3.4
		Average	12.54	12.54	11.26	9.48	9.48	5.58	5.58	7.8	6.72	5.36	5.36	5.36	4.52	4.4	1.56	0.1	-3
	Precipitation (P)	February	36	36	33	43	43	62	62	51	57	64	64	64	68	73	88	95	111
		March	31	31	33	43	43	69	69	58	65	74	74	74	78	78	96	105	126
		April	32	32	35	45	45	67	67	54	61	69	69	69	74	74	93	103	125
		May	21	21	24	31	31	48	48	39	44	50	50	50	54	54	68	77	95
		June	7	7	8	10	10	17	17	11	14	18	18	18	21	21	30	37	51
		Average	25.4	25.4	26.6	34.4	34.4	52.6	52.6	42.6	48.2	55	55	55	59	60	75	83.4	101.6

PART 2.3

METHODS: DATA ANALYSIS

2.3.1 SEQUENCING DATA ANALYSIS

For each region, genotypic data were analyzed to estimate genetic variability and DNA polymorphism and to generate phylogenetic trees on haplotype level (i.e. identical sequence across samples) and sample level (i.e. each sample represented by one sequence).

2.3.1.1 Genetic polymorphism

ARLEQUIN V3.5 (Excoffier & Lischer 2010) was used to estimate genetic variability indices and information as:

- Number of haplotypes *per* location (h/loc): number of haplotypes found in each location.
- Haplotype diversity (H_d): it measures the uniqueness of a particular haplotype in a given population.
- Nucleotide diversity (π %): it is the average number of nucleotide differences *per* site between any two DNA sequences chosen randomly from the sample population.
- Haplotype frequency: number of occurrences of a certain haplotype in a population divided by the total number of haplotypes in this population.
- Tajima's D neutrality test: it computes a standardized measure of the total number of segregating sites (these are DNA sites that are polymorphic) in the sampled DNA and the average number of mutations between pairs in the sample. If these two values only differ by as much as one could reasonably expect by chance, then the null hypothesis of neutrality cannot be rejected (Tajima 1989).

While DNASP V5 (Librado & Rozas 2009) was used to estimate the total haplotype diversity (H_d), total nucleotide diversity (π %) and Fu's neutrality test (synonym to Tajima's test).

Linkage disequilibrium (LD) analysis which tests the non-random association between sequenced regions was performed (i.e. combine

sequences of each region together for each sample) by using ARLEQUIN V3.5. The different alleles of the studied regions were codified at the haplotype level as standard data and the gametic phase was indicated as known. ITS1 and ITS2 were treated as different *loci*, because an initial inspection of the sequences suggested that there exists some recombination between the two parts of the ITS region. The significance level was set to p = 0.05.

2.3.1.2 Population structure and phylogeny

Phylogeny was performed using non-spatial descriptive methods. Only informative sites were used for phylogenetic reconstruction by excluding invariable sites among all sequences. Trees constructed at haplotype level, were inferred using the minimum spanning tree. The tree was estimated by ARLEQUIN V3.5 and visualized using HAPSTAR V0.7. Alternatively a distance-based approach, neighbor joining (NJ), was used in case that the secondary structures were considered using the web-based toolkit (ITS2 database server, http://its2.bioapps.biozentrum.uni-wuerzburg.de/; Koetschan *et al.* 2012). The indels were coded using SEQSTATE V1.4 (Müller 2005) by applying simple indel coding method (Simmons & Ochoterena 2000). Additionally the program was used to generate appropriate nexus files for PAUP and MRBAYES. Trees constructed at sample level were inferred using three sequence-based phylogenetic analyses approaches: maximum parsimony (MP) using PAUP*4b10 (Swofford 2002), maximum likelihood (ML) using MEGA V5 (Tamura *et al.* 2011) and Bayesian inference (BI) using MRBAYES V3.2 (Ronquist *et al*. 2012), the latter two need an evolution model test estimated apart by JMODELTEST V0.11 (Posada 2008) or MEGA V5. Tree construction was performed with a closely related species whenever it was possible to be used as an out-group. TRACER V1.5 (Rambaut & Drummond 2007) was used to evaluate BI method.

Gene flow (N_m) was estimated based on Φ_{ST} where “Nm = [(1 / Φ_{ST}) - 1] / 2”, calculated by GENALEX V6 (Peakall & Smouse 2006). ARLEQUIN V3.5 was used to generate pairwise genetic distance matrices and to test the population

differentiation measurement by performing the analysis of molecular variance (AMOVA). Table 2.9 summarizes all the analyses performed.

2.3.1.3 Software parameters

Parameters of each program were set up in order to proceed with the most suitable conditions for the data analysis of each region. Some of the programs parameters were depending on the results of the others. The parameters were set in the following order:

1- DNASP V5. Phylogenetic analysis program used for sequence collapsing (i.e. to generate haplotype list) and to estimate genetic diversity indices:
    a. The file with the aligned sequences was imported in the fasta format, then data information was given by the user:
        i. Nucleotide sequence format: nucleic acid type (DNA), genomic state (haploid), chromosomal location (autosome for nrDNA, mtDNA or cpDNA).
        ii. Sequence sets: samples were grouped in 17 datasets according to their sampling location.
        iii. Before analysis newly defined data was exported into ARLEQUIN file format, in form of haplotype list where invariable sites were removed and gaps were considered as fifth state.
        iv. Analysis procedure:
            1. DNA polymorphism/divergence data (total haplotype diversity and total nucleotide diversity).
            2. Fu's neutrality test
    b. The file with the aligned sequences of all regions *per* sample was imported in the fasta format.
2- ARLEQUIN V3.5. The manually formatted samples input file and the haplotype list input file (exported from DNASP V5) were loaded as two different projects. Settings were adjusted as follows:
    a. AMOVA analysis was selected.

b. A population differentiation test was performed with Markov chain of 100 000 and 10 000 dememorization steps, when significance level was 0.05.
c. LD analysis was performed.
d. Neutrality test under the infinite site model, Tajima's D was selected to be calculated from haplotypes inferred from distance matrix.
e. Diversity indices, molecular diversity options was selected:
    i. Standard diversity indices
    ii. Molecular diversity indices
f. From genetic structure, the AMOVA option was selected:
    i. Standard AMOVA
    ii. Compute minimum spanning tree
    iii. Compute distance matrix of pairwise differences

3- GENALEX V5 (Genetic Analysis in Excel). An add-in utility for Microsoft® EXCEL (Windows versions, EXCEL 97 or later) that contains tools for population genetics based on co-dominant, haploid and binary genetic data. The input file was prepared following the instruction manual to calculate the gene flow (N_m) for all locations.

4- SEQSTATE V1.4. A program for indel coding under various schemes. A fasta format alignment file had its indels (i.e. sites with gaps) codified using simple coding method (Simmons & Ochoterena 2000). The output file was in nexus format.

5- PAUP*4. It is a program used as a tool for inferring and interpreting phylogenetic trees, which can analyze molecular sequences, morphological data and other data types, using ML, MP and distance methods like the NJ method. The MP analyses were run with PAUP*4b10, using the following settings:
    a. Parsimony analysis was started (*Set criterion = parsimony MULTREES = yes steepest descent = no COLLAPSE = yes*).
    b. A bootstrap analysis (Felsenstein 1985) with 1 000 replicates was performed as (*Bootstrap Search = Heuristic np = 1 000 TreeFile = cp DNA.tre Brlens = yes*).

c. Generated trees were saved (*savetrees file = cpDNAFinal.tre savebootp = both from = 1 to = 1*).

6- JMODELTEST V0.11. The program is a tool to carry out statistical selection of best-fit models of nucleotide substitution. The Bayesian information criteria (BIC) was used to evaluate the best evolutionary model for each region, with the option to write PAUP command blocks to help in MRBAYES parameters setup.

7- MRBAYES V3.2. Software for Bayesian inference and model choice across a wide range of phylogenetic and evolutionary models. MRBAYES uses Markov Chain Monte Carlo (MCMC) methods to estimate the posterior distribution of model parameters. Based on the evolution model previously estimated, BI analysis was performed as follows:

a. To prepare the input nexus file with MRBAYES command block included with the following general format:

[*BEGIN MRBAYES;*
outgroup?;
lset Nst = ? rates = ?;
mcmcngen = 1 000 000 printfreq = 10 000 samplefreq = 100 nchains = 3 savebrlens = yes filename = ?;
sumt filename =? burnin = 100 000 contype = halfcompat;
*End;***]**

Where (?) depends on the DNA region.

b. To run the program and initiate the analysis using exe command followed by the nexus file of the targeted region for analysis.

8- TRACER V1.5. Tracer is a program for analyzing the trace files generated by Bayesian MCMC runs (that is, the continuous parameter values sampled from the chain). The program was used as follows:

a. Posterior results generated by MRBAYES were imported by drag-and-drop.

b. The evaluation based on two points:

i. To search for any problematic parameter indicated in red to identify the problem.

ii. To check the trace plot, in order to evaluate the MCMC stability.

9- MEGA V5. A phylogenetic program used for assembling data, evolutionary analysis and visualizing tree editor. Always based on Jukes-Cantor substitution model unless otherwise is mentioned. We used the program for:
    a. Computing the substitution model (a.k.a. evolution model), BIC was used for model evaluation.
    b. Constructing phylogenetic trees: ML based on the selected model, while all sites were included with bootstrap test of 1 000 replicates.
    c. Computing pairwise distance between generated haplotypes for each *locus* based on the maximum composite likelihood model, while gaps were not considered by default.

Table 2.9 Summary table of indices and analyses used for each category and the program used on sequencing data.

Category	Indices/analysis	Analysis program
Genetic diversity	Number of haplotypes *per* location	ARLEQUIN V3.5
	Nucleotide diversity (π%)	
	Haplotype diversity (H_d)	
	Observed sites with indels (OSI)	
	Haplotype frequencies	
	Tajima (D)	
	Fu (S)	DNASP V5
	Total nucleotide diversity (π%)	
	Total haplotype diversity (H_d)	
Phylogeny	Maximum parsimony	PAUP*4
	Evolution model	JMODELTEST V0.11/MEGA V5
	Maximum likelihood	MEGA V5
	Bayesian inference	MRBAYES V3.2
Population structure	Pairwise distance between haplotypes	MEGA V5
	Genetic distances (pairwise differences)	ARLEQUIN V3.5
	AMOVA	ARLEQUIN V3.5
	Gene flow (N_m)	GENALEX V6
Assessments test	Linkage disequilibrium (LD)	ARLEQUIN V3.5

2.3.1.4 Secondary structure

Whenever it was needed secondary structure prediction analysis was conducted. One objective was to confirm the loss of large fragments to investigate the mutational value of such loss. The other objective was to

improve the phylogenetic signal at the sample level. The following protocol was followed to fulfill the first objective:

1- The DNA mfold server (Zuker 2003) was used to input each of the generated haplotypes for DNA secondary structure prediction under the default parameters.
2- All generated structures were exported into the vienna format to VARNA V3.8 (Darty *et al.* 2009) for structure visualization.
3- Each generated structure for each haplotype was tested individually, focusing on the region where a fragment loss occurred (i.e. deletion of consecutive nucleotides found between multiple aligned sequences).
4- Results were exported to image format for editing and word processing.

The following protocol was used to fulfill the second objective:

1- To follow the latter protocol from step 1 to 3, and save each final structure for each haplotype/sample in a vienna format, resulting in a file in xfasta format (i.e. fasta file with structure in vienna format included for each sequence).
2- Import the xfasta file into ITS2 database server as user pool data. This database was designed for ITS2 secondary structure prediction and phylogenetic analysis only.
3- In the “Analyze data” section, we used the “Create alignment” feature by the “Sequence and structure” option to compare the secondary structures visually.
4- We used “Create tree” to generate a phylogenetic tree with the secondary structure considered, based on NJ method.

2.3.2 MICROSATELLITE DATA ANALYSIS

2.3.2.1 Fragment analysis

The methods of analyzing microsatellite data obtained from the fragment analysis are very sensitive to small bands, stutter bands, dimers and noisy background. In order to avoid any misleading data, we analyzed our data twice with different programs and methods (manual and automated methods). Both methods were compared and combined in some cases for fidelity:

2.3.2.1.1 Manual method

Trace files of each plate obtained from the fragment analysis service were loaded separately into a new project file generated by PEAKSCANNER™ Software (Applied Biosystems, U.S.A.). Trace files were analyzed in the following form:

1- New analysis method set. To set the peak detection parameters, certain modification from the default parameters of the program were set under the following conditions:
   a. Prior to the detection of peaks *per se*, a light smoothing of electropherograms had been set, in order to eliminate small secondary peaks inherent to technical background noise.
   b. The detection of peaks was achieved through a “sliding window” analysis that inspects electropherograms locally. Within the inspected region, PEAKSCANNER first creates a modeled version of the electropherogram by fitting a polynomial curve to the data. Peaks are detected according to this modeled signal, based on their absolute width. Therefore, the detection sensitivity is adjusted by modifying the width of the sliding window, the goodness of fit reachable by the polynomial curve and the minimal width above which a peak is recorded as present. Default parameters were used as a starting point, as they have been shown to provide reliable results (Holland *et al.* 2008; Herrmann *et al.* 2010). Parameters were set as 15 points

for the sliding window width, used a third degree polynomial curve and considered peaks that at least had two points of half-width.

c. Downstream to peak detection. PEAKSCANNER filters peaks according to their absolute fluorescence intensity, i.e. the peak height, measured in relative fluorescent units (rfu). Electropherograms were visually checked from blank and replicated samples, in order to adjust the fluorescence threshold to the upper limit of the technical background noise. 150 rfu was set as a minimal fluorescence for considering sample microsatellite peaks.

d. The customized analysis method was saved under the name “Microsatellite analysis method” in order to use it during electropherogram analysis.

2- Data import and preparation:

a. The electropherograms (*.fsa files) were imported into PEAKSCANNER project using the “Add files” button.

b. “GS500” was defined as the size standard and the “Microsatellite analysis method” as analysis method to be used for all samples included into the project (this information was for the first sample, then the columns “Size standards” and “Analysis method” were selected and by using the “ctrl D” keyboard shortcut. These settings were applied to all remaining samples).

3- Electropherogram analysis. The detection and sizing of peaks was processed using the “Analysis” menu, from the graphical interface. Once achieved, electropherograms can be visualized and compared among samples. Samples with a systematically low fluorescence or showing abnormal peaks were removed from the analysis and repeated apart.

4- Data Exportation. The PEAKSCANNER analysis ends with a simple export process in which the list of peaks detected throughout the complete set of analyzed samples is stored in a table. This is achieved using the “Export/Export Combined Table” menu, producing a text-

tabulated file containing the size, height, area and width of all detected peaks.

5- Allele detection and refining. The exported data were opened using Microsoft® EXCEL 2010, in order to proceed with the allele detection and comparison among samples, through the following steps:
    a. All columns were deleted except for (Dye/sample peak, Sample file name, Size, and Height) columns.
    b. A comparison was done between first raw (A1-A12) samples with their replicate (H1-H12) in order to determine the difference range, by subtracting the same peak molecular size (bp) of the first sample from its replicate. These values were used to manually detect bin ranges of the same peak.
    c. All peaks of the same bin were considered as one, stutter bands and very low intensity bands (less than 20 % of the highest peak) were excluded from the alignment.
    d. After allele determination all samples were compared visually based on their allele sizes for each *locus* (dye) alone.
    e. Finally an allelic table was generated manually for further analysis.

2.3.2.1.2 Automatic method

Trace files of each plate obtained from the fragment analysis service were sorted according to sample order. All trace files were imported into the GENEIOUS PRO 5.5.7 software where the microsatellite plugin (Drummond *et al.* 2011) was used to automatically analyze the generated peaks and generate allelic table in the following sequence:

1- Trace files in (.fsa) format were imported and stored in the GENEIOUS database.
2- Every single trace file was viewed and checked for amplification success.
3- Peaks were automatically called and the ladder was automatically matched to them.

4- The 250 bp peak on GS500 was automatically detected and removed, because it does not always migrate uniformly compared to the other peaks in the size standard (Klein *et al.* 2003).
5- *Locus* info was set in accordance with the SSR-*loci* information (*Locus* name, Repeat unit and Expected fragment size range).
6- Bins were predicted for every single peak with no specific scaling labeling method. In this way stutter bands effect was avoided.
7- The allele size graph was used to manage overall allele view.
8- Peak calls and sequences were exported in CSV format for further analysis.

2.3.2.2 Data analysis

As a co-dominant marker, SSR *loci* are read in an allelic form by all specialized programs. Allelic tables were generated in three forms depending on the analysis program used afterwards:

1- Letter/allele/*loci* form (i.e. sample 1, *locus* M2 its allelic form is written as A/A or B/B, which refer to a homozygous or A/B as a heterozygous).
2- Molecular size in numbers/allele/*loci* (i.e. sample 1, *locus* M2 its allelic form is written as 450/450 or 480/480, which refer to a homozygous or 450/480 as a heterozygous).
3- Repeat differences in numbers/allele/*loci* (i.e. sample 1, *locus* M2 its allelic form is written as 16/16 or 14/14, which refer to a homozygous or 16/14 as a heterozygous).

2.3.2.2.1 Genetic polymorphism

MICROSATELLITE TOOLKIT V3.1.1 (Park 2001) was used in order to estimate the genetic diversity levels through the measurement of various indices:

- Unbiased gene diversity (UH_Z): is the proportion of polymorphic loci across the genome measured in unequally sampled populations.

- Observed heterozygosity (OH_z): is the mean number of samples with polymorphic *loci*.
- Number of allele/*locus* (n_A): is the number of alleles *per* a studied *locus*.
- Effective number of alleles (n_e): is the number of alleles that reflect the genetic variability and is opposite to the homozygosity.
- Polymorphic index content (PIC): it is a measure of the allelic diversity.

2.3.2.2.2 Population structure

MULTILOCUS V1.3 (Agapow & Burt 2001) was used to assess the adequacy of the number of *loci* used. STRUCTURE V2.2 (Pritchard *et al.* 2000a; Hubisz *et al.* 2009) used to determine the population structure through a non-spatial Bayesian clustering method (NSBC). F_{ST} pairwise genetic distance and genetic differentiation through analysis of molecular variance (AMOVA) test were estimated using ARLEQUIN V3.5. Gene flow (N_m) was calculated by POPGENE32 V1.32 (Yeh & Boyle 1997). Table 2.10 shows a brief description of the performed analyses.

Table 2.10 Summary table of the population analysis category, indices and the program used to compute such indices based on microsatellite data.

Category	Indices	Analysis program
Genetic diversity	Unbiased gene diversity (UH_Z)	MICROSATELLITE TOOLKIT V3.1.1
	Observed Heterozygosity (OH_z)	
	Number of allele/*locus* (n_A)	
	Effective number of alleles (n_e)	
	Polymorphic index content (PIC)	
Population structure	Bayesian clustering: sub-population (K)	STRUCTURE V2.2
	Genetic differentiation (AMOVA)	ARLEQUIN V3.5
	Estimation of pairwise distance matrix	
	Gene flow (N_m)	POPGENE32 V1.32
Assessments tests	Adequacy of used *loci*	MULTILOCUS V1.3

2.3.2.2.3 Software parameters

Some programs needed a parameter set up in order to proceed with the most suitable conditions for SSR data type, the parameters set are sorted by their program usage as follows:

1- CONVERT. The program CONVERT facilitates the conversion of diploid genotypic data files into formats that can be directly read by commonly used population genetic computer programs, as in the current study: GENEPOP, ARLEQUIN, POPGENE, MICROSAT, PHYLIP and STUCTURE. The procedure was the following:
   a. The CONVERT input format was prepared starting from a Microsoft® EXCEL spread sheet containing the diploid genotypic data of all the SSR *loci*. The data were formatted and ordered as described in the program manual.
   b. After the input process, the data were converted and saved in separate folders for further analysis.

2- MULTILOCUS V1.3. It is a program that allows the calculation of various genotypic diversity indices, various linkage disequilibrium indices, and a measure of population differentiation from both dominant and co-dominant markers. The procedure was the following:
   a. The input file was formatted using Microsoft® EXCEL 2010 as described in the program manual, and imported to the program for further analysis.
   b. Pairwise linkage disequilibria (I) were calculated to test the independency of each *locus* from the others (only for the population structure inferring method).
   c. *Loci* number plotting against the mean genetic diversity was estimated, and visualized using Microsoft® EXCEL 2010.

3- STRUCTURE V2.2. The program implements a model-based clustering method for inferring population structure using molecular data consisting of *unlinked markers*. The method was introduced by Pritchard *et al.* (2000b) and extended in sequels by Falush *et al.* (2003, 2007). The method application is to detect the population structure, identifying distinct genetic groups, assigning samples to sub-

populations, and identifying migrants in admixed samples. The procedure was the following:

a. By using the independent *loci* data, a new project named "Microsatellite data" was created; stating the directory folder, number of samples and number of *loci*, while polyploidy was set to 2 (co-dominant marker). One extra row and two extra columns identification were set as "Marker names, Label (sample number/name) and PopData", respectively.
b. Parameters were set as following burn-in of 10 000, a run length of 100 000, and admixture ancestry model (each sample draws some fraction of its genome from each of the K populations) with correlated allele frequencies.
c. Triple simulations (Run) were performed by setting the number of sub-populations (K) from 1 to 10. The program calculated the estimated natural logarithm probability score "estimated ln Prob (K)", "mean ln Prob (K)" and "variance of ln likelihood" for each run.
d. The subpopulation number K selection criterion is to detect the most probable value of K.
e. When detected, the highest two K in ln Prob (K) values are to be selected to perform a new run *per* K (five times) in order to ensure the ln Prob (K) values.
f. The bar plot of the highest Ln Prob (K) value was then exported to image format for processing.

4- ARLEQUIN V3.5:
   a. An input file was loaded into the program where data were scored for the number of repeats of each allele/*loci*.
   b. Analysis of molecular variance (AMOVA) test was performed using the default AMOVA settings.
   c. The pairwise genetic distance was estimated based on genetic distances F_{ST}.
   d. Results were copied and reformatted for word processing.

5- MICROSATELLITE Toolkit V1.3. It is an add-in utility for Microsoft® EXCEL (Windows versions, EXCEL 97 or later) that contains tools for

population genetics based on microsatellites data. The procedure was the following:

a. Input files were formatted in Microsoft® EXCEL 2010 as shown in the program manual and then the add-in was activated.
b. Indices were calculated and saved automatically in two ways, populations (separated) and (combined); while the error check ability was set on, and *locus* size ranged from 100 to 500.

6- POPGENE32 V1.32. It is software for the analysis of genetic variation among and within populations using co-dominant and dominant markers. The procedure was the following:

a. Input files generated by CONVERT were imported as co-dominant marker data.
b. The co-dominant menu of the indices estimation and genetic variation tests was selected. Data were analyzed as a single population (i.e. each location alone) and as a multiple population (i.e. all the locations were analyzed together).
c. Gene flow (N_m) estimation was performed based on wright's fixation index F_{ST} (i.e. an index which measures the genetic correlation between pairs of genes sampled within a location relative to pairs of genes sampled within the overall set of locations), where "Nm = [(1 / F_{ST}) - 1] / 2".

2.3.3 AFLP DATA ANALYSIS

2.3.3.1 Fragment analysis

AFLP chromatogram data obtained from the fragment analysis method are very sensitive to factors like small bands, stutter bands, dimers, and noisy background. To minimize the influence of such factors, an automated AFLP scoring was performed using both programs, PEAKSCANNER for peak analysis and RAWGENO V2 for automated scoring (Arrigo *et al.* 2009).

2.3.3.1.1 Peak analysis

Trace files of each plate obtained from the fragment analysis service were loaded separately into a new project file generated by PEAKSCANNER. Trace files were analyzed in the following form:

1- New analysis method set. The analysis method was set as described for microsatellite analysis, except for:
   a. The PEAKSCANNER filters for the absolute fluorescence intensity was 150 rfu as a minimal fluorescence for considering sample AFLP peaks.
   b. The customized analysis method was saved under the name “AFLP analysis method” in order to use it during electropherogram analysis.
2- Data import and preparation was as described before (point 2.3.2.1.1).
3- Electropherograms analysis was performed as described before (point 2.3.2.1.1).
4- Data were exported to a text-tabulated file containing the size, height, area and width of all detected peaks.

2.3.3.1.2 Automated AFLP scoring

The RAWGENO V2 program uses the R PACKAGE V4 (Casgrain & Legendre 2001) as a platform, and is able to automatically detect and score AFLP

peaks imported from PEAKSCANNER. It includes an automatic bin detection function and some statistical tests in order to ensure the obtained scoring fidelity and to reach the maximum accuracy.

The process had the following order:

1- Importing data into RAWGENO V2. During importation, RAWGENO V2 handles a single dye color at a time, which is user-specified and considers the “dye” parameter with the following values: "B" (blue; FAM), "G" (green; HEX), "Y" (yellow; NED = CY3), “R” (red; ROX) or “O” (orange; LIZ). In the case of multiplexing of PCR products, each dye had to be analyzed separately. Datasets obtained from several dyes were merged *a posteriori* in a final binary table (Fig. 2.2).

2- Filtering of low quality samples. The low quality samples were detected to prevent false-absences in the final dataset, through a check test they pop up automatically after data importation. That shows the variability in the number of peaks detected *per* sample which is used as a proxy of AFLP reactions quality. All low quality samples, when found, were discarded, re-amplified and sent for the fragment analysis service again.

3- Scoring. As automatic program, RAWGENO fully scores the AFLP peaks through two steps: binning followed by filtering of the predicted bins. The procedure was the following:

   a. Binning. RAWGENO uses a binning algorithm relying on the size of AFLP peaks over all samples included in the project and defines bins in a way that respects the two following conditions:

      i. Maximal bin width. This prevents the definition of too large bins that could lead to homoplasy (i.e. erroneously assigning non-homologous AFLP peaks within the same bin). This limit was set using the MaxBin parameter to 1.5.

      ii. Minimum bin width. This prevents the assignment of more than one peak from the same sample within the same bin (i.e. “technical homoplasy”). This limit was set using the MinBin parameter to 1.

b. Filtering. Once defined, bins can be filtered according to their properties and/or quality. Three filters were used as defined by the program manual as following:
    i. The size filter. The binning of peaks included in the range of the size ladder was limited, peaks with small sizes (i.e. smaller than 150 bp, RMIN = 150) according to Vekemans *et al.* (2002) were discarded, where normally dense-band regions show up.
    ii. Fluorescence average filter. Based on the assumption that bins with a high average fluorescence retrieve a more consistent signal than bins with a low fluorescence, some bins were eliminated according to their average fluorescence. This filter was set by 150 as low fluorescence bins (stdRFU) and 1 for low frequency bins (Nblnds) default.
    iii. The reproducibility filter. It evaluates bin quality according to their robustness across AFLP reactions, by relying on replicated samples. This filter assumes that replicated samples were selected randomly from the original dataset, in a way to scan better the genetic diversity. RAWGENO identifies replicated samples using their names, and for each bin, RAWGENO compares original to replicated samples and calculates the percentage of original-replicated pairs for which the AFLP signal was successfully reproduced. Bins where reproducibility cannot reach a satisfactory rate were eliminated from the final dataset. This filter was set on with 80 % of reproducibility, while the untested bins elimination option was on.

4- Quality checking. Scoring quality checking tests were held to check both the binning and samples quality:
    a. Binning quality:
        i. Manual check. As binning is an automated and straightforward analysis, all bins were preferred to be

revised manually by interactive method through the RAWGENO visualization option for manually editing the binning by adding, removing or modifying the width and position of bins. Based on several help-to-decision statistics included in the program such as the average size and the number of presences associated *per* bin, decisions were made independently for each bin.

ii. After binning, some quality statistics were used and demonstrated visually according to the "bin size" to ensure the final bin quality. These were:

1. Bin width, the difference between the largest and shortest AFLP bands included into the focal bin.
2. Technical homoplasy, the average number of peaks belonging to same sample, included into the same bin.
3. Size homoplasy, according to Vekemans *et al*. (2002), size homoplasy can be detected by measuring and testing the linear correlation existing between the size of bins and their frequency.
4. Peak intensity, the fluorescence of AFLP peaks.

b. Sample checking. RAWGENO was used to display the exploring scoring results through two ways (menu "RAWGENO/Quality Check/Samples Checking"). Both ways were handled through a graphical user interface for sorting and selecting samples to be visualized. One is that the binary matrix is directly visualized using a heat map, showing samples sorted according to their genetic relatedness. Alternatively, samples were examined with a principal co-ordinates analysis. Both ways were used to compare AFLP results with either quality statistics (i.e. number of AFLP peaks per sample, mean and variance in fluorescence intensity and outlier detection index).

5- Result exporting. For each dye “Primer combination” a final transposed binary tables was generated (i.e. samples = rows) constantly and saved in a separate archive.

6- Merging data from several AFLP markers. Every three transposed binary tables were merged into a single matrix using the RAWGENO import menu options (menu Files / Import / Merge several datasets (binary tables)) then saved and exported as one transposed binary table (Fig 2.2). Since the *loci* names and numbers were different for each plate (triplex), therefore all four plates were merged afterwards manually using Microsoft® EXCEL 2010, and saved in the same transposed binary table form.

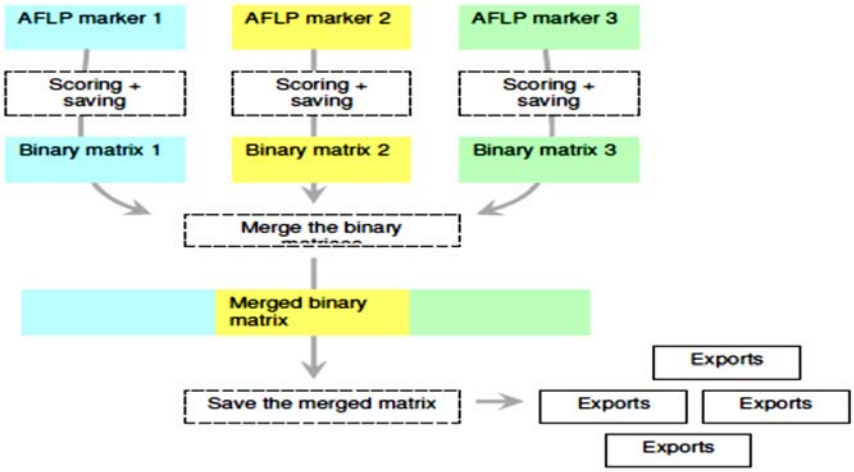

Figure 2.2 Merging several datasets in AFLP fragment analysis. Strategy to merge several datasets scheme represented from RAWGENO manual file.

2.3.3.2 Data analysis

The analysis of the AFLP data was performed in two known approaches. The first is based on the band-binary criterion (i.e. codifying the detected bands to 1 when presence and 0 when absent) and the other is based on the allele frequency (i.e. number of a band presence relatively to the number of all individuals). Each has its advantages and disadvantages, therefore, both forms were combined to obtain the maximum number of valuable indices according to Bonin *et al.* (2007) methodology.

2.3.3.2.1 Genetic polymorphism

Genetic variability was demonstrated by measuring valuable indices that were measured using FAMD (Schlüter & Harris 2006) and ARLEQUIN V3.5:

- Number of polymorphic bands (PB): number of *loci* inconstantly scored through all samples.
- Percentage of polymorphic *loci* (PPL): percentage of PB.
- Number polymorphic bands at 5 % level (PB at 5 %): number of polymorphic *loci* with frequency > 5 % and < 95 %.
- Proportion of polymorphic *loci* at 5 % level (PLP at 5 %): number polymorphic bands at 5 % divided by PB.
- Number of private bands (PrB): number of polymorphic *loci* scored only for a specific location.
- Number of fixed private bands (FB): number of private bands scored for all the samples of a specific location.
- Expected heterozygosity under Hardy-Weinberg (H_e): mean value of all the expected level of heterozygosity found for all *loci* in a sample.
- Effective number of alleles (n_e)

2.3.3.2.2 Population structure and phylogeny

MULTILOCUS V1.3 was used to assess the adequacy of the number of *loci* used. The population structure was revealed through a non-spatial Bayesian clustering method (NSBC) by using STRUCTURE V2.2. AFLP-SURV (Vekemans *et al.* 2002) was used to calculate gene flow (N_m) and F_{ST} pairwise genetic distance matrix when the inbreeding value (F_{IS}) was set to 1 (complete homogeneity was assumed due to the lack of information about the ploidy level of the samples included in this study). ARLEQUIN V3.5 was used to test the population differentiation measurement by performing the analysis of molecular variance (AMOVA). HAPSTAR V0.7 (Teacher & Griffiths 2011) was used to draw the haplotype network generated by ARLEQUIN V3.5. Table 2.11 shows a brief description of the performed analysis.

Table 2.11 Summary table of the population analysis category, indices, approach in which they are based and the program used to compute such indices based on AFLP data.

Category	Indices	Approach	Analysis program
Genetic diversity	n_e, H_e	Band-based	ARLEQUIN V3.5
	PPL, PB, PrB, FB	Band-based	FAMD
Population structure	Genetic distance matrix	Band-based	ARLEQUIN V3.5
		Allele-based	AFLP-SURV
	NSBC	Allele-based	STRUCTURE V2.2
	AMOVA	Band-based	AFLP-SURV
	N_m	Band-based	
Assessments tests	Adequacy of used *loci*	Band-based	MULTILOCUS V1.3

2.3.3.2.3 Software parameters

Some programs needed a parameter set up in order to proceed with the most suitable conditions for AFLP "band-based" or "allele frequency" data type. In no particular order the parameters were set as follows:

1- FAMD. Fingerprint analysis with missing data software, specially used with fingerprint genetic data. The program was used:
   a. To estimate frequencies *per* samples.
   b. To count *loci*, mean number of bands *per* sample, polymorphic bands and fixed bands.
   c. To remove *loci* with frequency < 5 % and > 95 % (at 5 %) and to calculate polymorphic bands, portion of polymorphic bands and fixed bands, at 5 % level.
   d. To export files into GENEPOP input format for further analysis.

2- MULTILOCUS V1.3. The procedure was the following:
   a. The input file was formatted using Microsoft® EXCEL 2010 as described in the program manual, and imported to the program for further analysis.
   b. Plot *loci* at 5 % against their calculated genetic diversity, and visualized by Microsoft® EXCEL 2010.

3- AFLP-SURV. It estimates genetic diversity and population genetic structure from population samples analyzed with AFLP or RAPD methods and computes genetic distance matrices between populations

by using Lynch & Milligan (1994) treatment, which uses the average expected heterozygosity of the marker *loci*, or Nei's gene diversity, as a measure of genetic diversity. The procedure was the following:

a. The process started by estimating allelic frequencies at each marker *locus* in each population based on fragment frequency.
b. All default parameters were used without modification as described in the program manual to produce matrices of pairwise genetic distances between populations (with bootstraps = 100) and of pairwise relatedness coefficients between samples, while automatically various tests of significance based on random permutations were performed.
c. The genetic distance matrix (F_{ST} based) was reformatted and exported for further analysis.

4- STRUCTURE V2.2. A new project named “AFLP band-based” was created, while the same parameters were set and used as described before (point 2.3.2.2.3).

5- ARLEQUIN V3.5. The procedure was the following:

a. The input file was created manually and loaded as RFLP haploid data into the program, where data were scored in a binary form.
b. An analysis of molecular variance (AMOVA) test was performed using the default AMOVA settings while the minimum spanning tree option was set on.
c. A minimum spanning tree between haplotypes was converted into a minimum spanning network when alternative connections were included.

7- HAPSTAR V0.7. Program used to automatically ordinate and visualize the minimum spanning network, with the option to control the speed of sorting and the size of the network. Finally the results were exported as image format for word processing.

2.3.4 CLIMATIC DATA ANALYSIS

This analysis aimed to study the possible relation between pairwise genetic distances found at locations level with other factors associated to each location, like climatic variations, geographical distance or altitudinal gradient. We used two methods to find and measure such relations, first through a genome scan approach and the other through correlation tests.

2.3.4.1 Genome scan and *loci* under selection

This procedure identifies *loci* under natural selection or outliers (i.e. *loci* in which exhibit higher or lower F_{ST} values than the great majority of neutral markers) by searching for the genetic signatures of selection, screening 1 584 AFLP bands. This test was used with the aim to identify specific genome regions or genes (i.e. AFLP scored bands) associated with adaption to bioclimatic belts through the altitudinal gradient of *Funaria hygrometrica* populations in the Sierra Nevada Mountains. We used MCHEZA (Antao & Beaumont 2011) to detect positive outlier under the following scheme:

1- A GENEPOP input format file previously generated by FAMD was used as MCHEZA input file twice: a) as 17 sets based on the location numbers, b) as 4 sets based on the altitudinal groups (see pp. 56).
2- Automatically the program calculates the attempt (neutral) F_{ST} value.
3- Under the default parameters, the program was run five times with 100 000 simulations at 99.5 % confidence limit.
4- Positive outlier *loci* extracted from each run were compared with each other's. *Loci* showed to be outlier in less than 3 runs were excluded.
5- A new data sheet of the 84 samples with only the outlier *loci* was generated for isolation by adaptation tests.

AMOVA test was performed using ARLEQUIN V3.5 for AFLP outliers *loci* dataset, and for AFLP neutral *loci* dataset (all AFLP *loci* except outliers) to compare the variation partition within locations. Outliers expected to increase F_{ST} value and decrease the variation detected within locations. The program was used as previously described (point 2.3.3.2.3).

2.3.4.2 Correlation tests

The population genetic structure of *Funaria hygrometrica* in Sierra Nevada was subjected to correlation tests with, 1) spatial patterning of climatic and bioclimatic variation (a.k.a. isolation by adaptation, IBA), and 2) geographic distance (a.k.a. isolation by distance, IBD). Simple Mantel tests (to measure the association between two matrices) were performed by plotting all possible combinations. The genetic distance and log (genetic distance) matrices of each marker (including the AFLP positive outlier *loci*) were tested against the Euclidean distance and the log (Euclidean distance) matrices of the environmental variables (geographic, climatic and bioclimatic) in order to find the highest correlation coefficients. Data and log data were used to find which were the most appropriate to represent a better correlation using Mantel test (Bohonak 2002).

The Mantel test was performed using GENALEX V6. Input files (genetic data and geographic data) were formatted in Microsoft® EXCEL 2010 as shown in the program manual and then the add-in was activated. The analysis was performed for each genetic marker as follows:

a. Geographic coordinates were loaded and a distance matrix between sampling locations was estimated. The Mantel test was then performed using the default preset values and parameters (999 permutation steps) between the genetic distance and geographic distance.
b. Climatic or bioclimatic Euclidean distance or log (Euclidean distance) matrices generated by PCO3 were loaded in place of the geographic distance matrix and tested with a Mantel test against the preloaded genetic distance matrix using the default preset values and parameters (999 permutation steps).
c. Genetic distance matrix of each genetic marker was tested against all other genetic markers.

All results were sorted in ascending form to find the highest association for each tested marker, and plotted for further illustration using Microsoft® EXCEL 2010.

CHAPTER 3

RESULTS

PART 3.1

Genetic Diversity

&

Population Analysis

3.1.1 NUCLEAR DNA – INTERNAL TRANSCRIBED SPACER 1 (ITS1)

3.1.1.1 Genetic variation and DNA polymorphism

The internal transcribed spacer 1 (ITS1) sequences for 84 *Funaria hygrometrica* samples from 17 locations ranged from 244 to 255 bp in size, separated into two groups of 70 occurrences with 244 to 249 bp and 14 occurrences with 254 to 255 bp in size (Table 3.1).

A total of 22 haplotypes (GenBank accessions: JX985532 – JX985553) were defined by 75 variable sites with 19 mutations in 18 polymorphic (segregating) sites of which three sites were singletons (two-variants) and 15 were parsimony informative, while 57 gap-residues were inserted to keep the sequences aligned forming 11 indels (Table 3.2).

The number of haplotypes *per* location ranged from one to six. Locations 8 and 10 had the highest number of different haplotypes, followed by locations 5, 9 and 12 with five different haplotypes, location 11 had four haplotypes, while location 13 had three haplotypes. Locations 3, 14, 15 and 16 had two haplotypes. The rest of locations had only one haplotype sequence (Table 3.2).

When gaps were excluded, haplotype diversity was 0.52 and nucleotide diversity was 0.63 % for all samples combined together. Locations 3, 14 and 15 showed the highest levels of haplotype diversity (1.00) and nucleotide diversity of 0.07, 0.02 and 0.71 %, respectively. Both locations 1 and 17 contained a single sample and therefore scored haplotype diversity of 1.00 but with no nucleotide diversity. Locations 2, 4, 6 and 7 showed a haplotype and nucleotide diversity of 0.00. (Table 3.2).

A significant negative Tajima's D test value ($D = -1.83$, p-value < 0.05) was obtained, which rejects the null hypothesis of neutral evolution and demographic equilibrium of the ITS1 spacer for all the locations combined together. The un-neutral state of ITS1 was confirmed by Fu's test of a significant negative value ($Fs = -7.90$, p-value = 0.00).

Table 3.1 ITS1 data obtained from *Funaria hygrometrica*. Locations, samples, haplotypes and size in bp of ITS1 spacer, are shown.

Location	Sample	Haplotype	Size (bp)	Location	Sample	Haplotype	Size (bp)
1.1	1	1	249	10.1	43	3	248
2.1	2	2	244	10.2	44	6	249
2.2	3	2	244	10.3	45	3	248
2.3	4	2	244	10.4	46	5	248
3.1	5	3	248	10.5	47	15	248
3.2	6	4	249	10.6	48	3	248
4.1	7	3	248	10.7	49	1	249
4.2	8	3	248	10.8	50	3	248
4.3	9	3	248	10.9	51	17	248
5.1	10	5	248	10.10	52	3	248
5.2	11	6	249	11.1	53	3	248
5.3	12	7	249	11.2	54	3	248
5.4	13	8	248	11.3	55	15	248
5.5	14	8	248	11.4	56	3	248
5.6	15	3	248	11.5	57	3	248
5.7	16	8	248	11.6	58	3	248
5.8	17	3	248	11.7	59	3	248
6.1	18	3	248	11.8	60	9	254
6.2	19	3	248	11.9	61	18	248
6.3	20	3	248	11.10	62	15	248
6.4	21	3	248	12.1	63	20	249
7.1	22	9	254	12.2	64	5	248
7.2	23	9	254	12.3	65	3	248
7.3	24	9	254	12.4	66	9	254
8.1	25	10	248	12.5	67	3	248
8.2	26	3	248	12.6	68	9	254
8.3	27	11	249	12.7	69	3	248
8.4	28	12	247	12.8	70	19	255
8.5	29	3	248	13.1	71	3	248
8.6	30	3	248	13.2	72	3	248
8.7	31	6	249	13.3	73	9	254
8.8	32	13	248	13.4	74	3	248
8.9	33	3	248	13.5	75	21	248
9.1	34	3	248	13.6	76	9	254
9.2	35	14	249	14.1	77	3	248
9.3	36	3	248	14.2	78	6	249
9.4	37	9	254	15.1	79	3	248
9.5	38	15	248	15.2	80	22	254
9.6	39	3	248	16.1	81	9	254
9.7	40	16	248	16.2	82	9	254
9.8	41	3	248	16.3	83	15	248
9.9	42	14	249	17.1	84	9	254

Table 3.2 DNA polymorphism of ITS1 sequences. Sample locations, sample size (N), polymorphic *loci* (PL), number of haplotypes *per* location (h/loc), haplotype diversity (H_d), standard deviation of haplotype diversity (H_d s.d.), nucleotide diversity (π %) and standard deviation of nucleotide diversity (π % s.d.), are shown.

Location	N	PL*	h/loc	H_d	H_d s.d.	π %	π % s.d.
1	1	0	1	1.00	0.00	0.00	0.00
2	3	0	1	0.00	0.00	0.00	0.00
3	2	0	2	1.00	0.50	0.07	0.08
4	3	0	1	0.00	0.00	0.00	0.00
5	8	2	5	0.86	0.11	0.04	0.03
6	4	0	1	0.00	0.00	0.00	0.00
7	3	0	1	0.00	0.00	0.00	0.00
8	9	8	6	0.83	0.13	0.09	0.06
9	9	3	5	0.81	0.12	0.17	0.10
10	10	3	6	0.78	0.14	0.03	0.02
11	10	2	4	0.64	0.15	0.15	0.08
12	8	1	5	0.86	0.11	0.37	0.21
13	6	2	3	0.73	0.16	0.38	0.10
14	2	0	2	1.00	0.50	0.02	0.03
15	2	1	2	1.00	0.50	0.71	0.72
16	3	1	2	0.67	0.31	0.47	0.36
17	1	0	1	1.00	0.00	0.00	0.00
All samples	84	18	22	0.52*	-	0.63*	-

***** Calculated when indels were excluded

Haplotype 3 was the most abundant (37 samples) and occupied a wide geographic distribution (12 locations), followed by haplotype 9 with 12 samples and present in 7 locations. Haplotypes 1, 2, 5, 6, 8, 14 and 15 were less abundant with 2, 3, 3, 4, 3, 2 and 5 samples and present in 2, 1, 3, 4, 1, 1 and 4 locations, respectively. The rest of haplotypes occurred only once. Only locations 9, 11, 12 and 13 shared the two most abundant haplotypes (3 and 9) (Table 3.3).

Table 3.3 Frequencies of 22 ITS1 haplotypes in 17 locations and the number of samples (S) counted for each haplotype. The frequency of each haplotype *per* location is equal to the number of samples of the same haplotype divided by the total number of samples found in that location.

Haplotype	Location																	S
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	
1	1.00	-	-	-	-	-	-	-	-	0.10	-	-	-	-	-	-	-	2
2	-	1.00	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	3
3	-	-	0.50	1.00	0.25	1.00	-	0.44	0.44	0.50	0.60	0.38	0.50	0.50	0.50	-	-	37
4	-	-	0.50	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
5	-	-	-	-	0.13	-	-	-	-	0.10	-	0.13	-	-	-	-	-	3
6	-	-	-	-	0.13	-	-	0.11	-	0.10	-	-	-	0.50	-	-	-	4
7	-	-	-	-	0.13	-	-	-	-	-	-	-	-	-	-	-	-	1
8	-	-	-	-	0.38	-	-	-	-	-	-	-	-	-	-	-	-	3
9	-	-	-	-	-	-	1.00	-	0.11	-	0.10	0.25	0.33	-	-	0.67	1.00	12
10	-	-	-	-	-	-	-	0.11	-	-	-	-	-	-	-	-	-	1
11	-	-	-	-	-	-	-	0.11	-	-	-	-	-	-	-	-	-	1
12	-	-	-	-	-	-	-	0.11	-	-	-	-	-	-	-	-	-	1
13	-	-	-	-	-	-	-	0.11	-	-	-	-	-	-	-	-	-	1
14	-	-	-	-	-	-	-	-	0.22	-	-	-	-	-	-	-	-	2
15	-	-	-	-	-	-	-	-	0.11	0.10	0.20	-	-	-	-	0.33	-	5
16	-	-	-	-	-	-	-	-	0.11	-	-	-	-	-	-	-	-	1
17	-	-	-	-	-	-	-	-	-	0.10	-	-	-	-	-	-	-	1
18	-	-	-	-	-	-	-	-	-	-	0.10	-	-	-	-	-	-	1
19	-	-	-	-	-	-	-	-	-	-	-	0.13	-	-	-	-	-	1
20	-	-	-	-	-	-	-	-	-	-	-	0.13	-	-	-	-	-	1
21	-	-	-	-	-	-	-	-	-	-	-	-	0.17	-	-	-	-	1
22	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.50	-	-	1

3.1.1.2 Population structure and phylogeny

3.1.2.1 Haplotype level

When indels are counted as fifth state, a minimum spanning tree (Fig. 3.1) showed a separation of the haplotypes into three groups. Group A, present in 67 samples, included haplotype 3 (the most abundant haplotype, present in 37 samples), together with the rest of haplotypes with the exception to haplotypes 2, 9, 19 and 22, and is represented in blue color. Haplotype 10 was found at a distance from the main group of eight mutational steps. Group B, present in 14 samples, included haplotype 9 (the second most abundant haplotype, present in 12 samples), together with haplotypes 19 and 22 (one sample each), and is marked in red color. The difference between group A and group B was estimated as 48 mutational steps. Haplotype 2 was found on 21 mutational steps from group A and was considered a group with a unique haplotype (group C), present in three samples, and is marked in green.

The main differences between the three haplotype groups consisted in: a) a fragment of 10 nucleotides found in groups A and C while it is missing in group B (indel-1); b) a fragment of 5 nucleotides found in group A which is absent in groups B and C (indel-2); c) a fragment of 26 nucleotides found in group B, and absent in groups A and C (indel-3). Group C generally was in accordance with group A in respect to these mutational events. Small divergences were present in each haplotype group (Fig. 3.1).

After excluding the gaps, a pairwise distance matrix between 22 haplotypes was generated with MEGA based on the maximum composite likelihood model. The minimum genetic distance (0.00 %) was scored from the comparison between haplotypes 3 – (4, 6, 7, 9, 19 and 20), 4 – (6, 7, 9, 19 and 20), 6 – (7, 9, 19 and 20), 7 – (9, 19 and 20), 9 – (18, 19 and 20), 18 – 20 and 19 – 20. The differences between these haplotypes consist only in indels and were therefore not taken into account in the genetic distance calculations. The maximum genetic distance value (6.99 %) was observed comparing haplotypes 2 – 10 (Table 3.4).

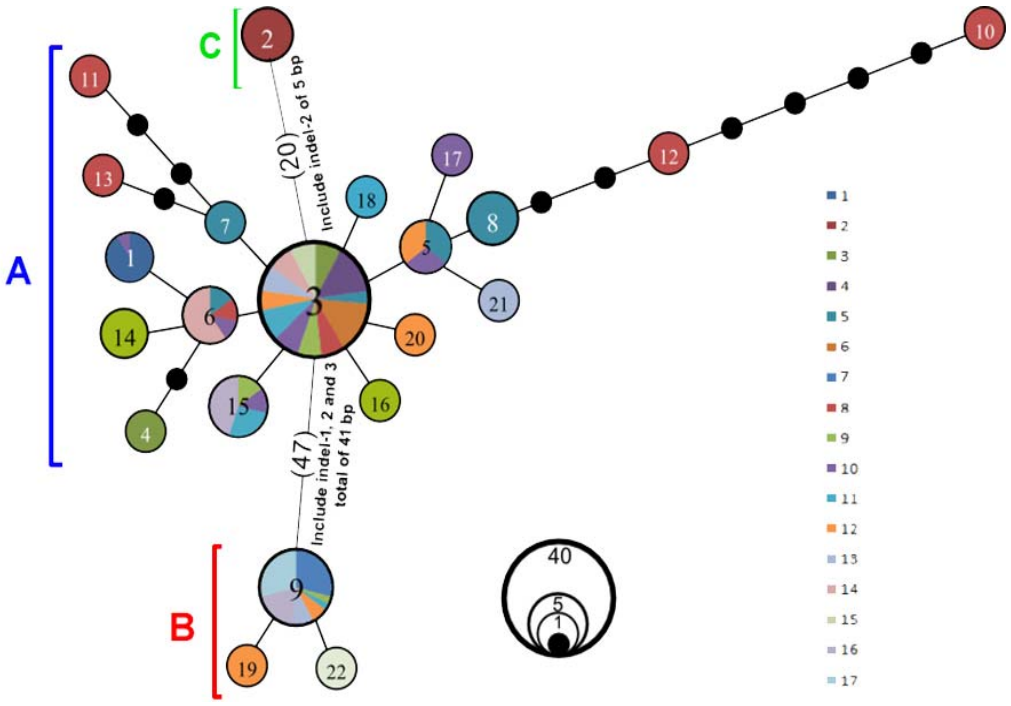

Figure 3.1 Minimum spanning tree of 22 ITS1 haplotypes from 84 samples in 17 locations. The size of the circles is proportional to the haplotype abundance. Small black circles are hypothetical intermediate haplotypes not found in our study. Group A (blue), B (red) and C (green) are illustrated in the figure, while the number of hypothetical intermediate samples are shown between brackets with indication of the indels portion to the demonstrated distance.

3.1.1.2.2 Sample level

The phylogenetic tree of ITS1 sequences generated using BI was confirmed by both MP and ML methods (Fig. 3.2). The tree was rooted with *Funaria microstoma* Bruch & Schimp. (GenBank: JN089175.1).

The basal part of the tree is not well resolved. Two main clades were clearly distinguished, but because of the uncertainty of the root part of the tree, one cannot be sure that these clades are monophyletic or paraphyletic.

Clade A supported with 0.70 of BI posterior probability and 63.7 bootstrap support of MP method, was divided into two sub-clades. Sub-clade

A.I was formed by Sierra Nevada samples of haplotypes belonging to group A of the minimum spanning tree (Fig. 3.1) along with samples from Italy (Sicily), Greece (sample number 1) and England. Sub-clade A.II was formed by Sierra Nevada samples of haplotype 2 (group C) along with German accessions (GenBank: X80212.2 and X74114.1), Italy, U.S.A. samples from Connecticut (CN 1, 2, 3 and 4), North Carolina (NC 2) West Siberia and *Funaria flavicans* Michx. (GenBank: JN089173.1).

Clade B supported with 0.86 of BI posterior probability and 68.2 bootstrap support of MP method, was divided into two sub-clades. Sub-clade B.I was formed by samples from Sierra Nevada of haplotypes belonging to group B of the minimum spanning tree along with one U.S.A. sample (GenBank: JN089174.1). Sub-clade B.II was formed by samples from Turkey (samples number 1 and 2) and one U.S.A. sample from California (CA).

3.1.1.2.3 Location level

When locations 1 and 17 were excluded, gene flow (N_m) of 0.85 based on Phi-$_{ST}$ (Φ_{ST}) was calculated by GENALEX software.

An AMOVA test was estimated by GENALEX software based on the pairwise differences, to test the genetic structure by comparing results from various groupings and choosing the structure that maximizes the among locations variance. Φ_{ST} was equal to 0.37, partitioned into a major genetic variation originated within locations, accounting for 63 % of the total variations, while 37 % of the genetic variation occurred among locations (Table 3.5).

Table 3.4 Pairwise distance matrix between 22 ITS1 haplotypes (H). Values are given in percentages, while minimum and maximum values are written in bold.

H	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1																						
2	4.29																					
3	0.40	3.85																				
4	0.40	3.87	**0.00**																			
5	0.81	4.29	0.40	0.41																		
6	0.40	3.85	**0.00**	**0.00**	0.40																	
7	0.40	3.85	**0.00**	**0.00**	0.40	**0.00**																
8	1.22	4.73	0.81	0.81	0.40	0.81	0.81															
9	0.44	3.62	**0.00**	**0.00**	0.44	**0.00**	**0.00**	0.89														
10	3.30	**6.99**	3.30	3.31	2.88	3.30	3.30	2.46	3.15													
11	0.81	5.18	1.22	1.22	1.63	1.22	1.22	1.22	1.33	2.46												
12	2.05	5.65	1.64	1.64	1.22	1.64	1.64	0.81	1.79	1.64	1.23											
13	0.81	4.31	0.41	0.40	0.81	0.41	0.40	1.22	0.44	2.89	0.81	1.22										
14	0.81	4.29	0.40	0.40	0.81	0.40	0.40	1.22	0.44	3.72	1.63	2.05	0.81									
15	0.81	4.29	0.40	0.41	0.81	0.40	0.40	1.22	0.44	3.72	1.63	2.05	0.81	0.81								
16	0.81	4.29	0.40	0.41	0.81	0.40	0.40	0.40	0.44	2.88	0.81	1.22	0.81	0.81	0.81							
17	1.22	4.73	0.81	0.81	0.40	0.81	0.81	0.81	0.89	3.30	2.05	1.64	1.22	1.22	0.40	1.22						
18	0.81	3.85	0.40	0.41	0.81	0.40	0.40	1.22	**0.00**	2.88	1.63	2.05	0.81	0.81	0.81	0.81	1.22					
19	0.40	3.85	**0.00**	**0.00**	0.40	**0.00**	**0.00**	0.81	**0.00**	3.30	1.22	1.64	0.41	0.40	0.40	0.40	0.81	0.40				
20	0.44	3.61	**0.00**	**0.00**	0.44	**0.00**	**0.00**	0.89	**0.00**	3.15	1.33	1.79	0.44	0.44	0.44	0.44	0.89	**0.00**	**0.00**			
21	0.81	4.73	0.81	0.81	0.40	0.81	0.81	0.81	0.89	2.46	1.63	1.64	1.22	1.22	1.22	1.22	0.81	1.22	0.81	0.89		
22	0.89	4.09	0.44	0.44	0.89	0.44	0.44	1.33	0.39	3.61	1.78	2.24	0.89	0.89	0.89	0.89	1.33	0.44	0.44	0.39	1.33	

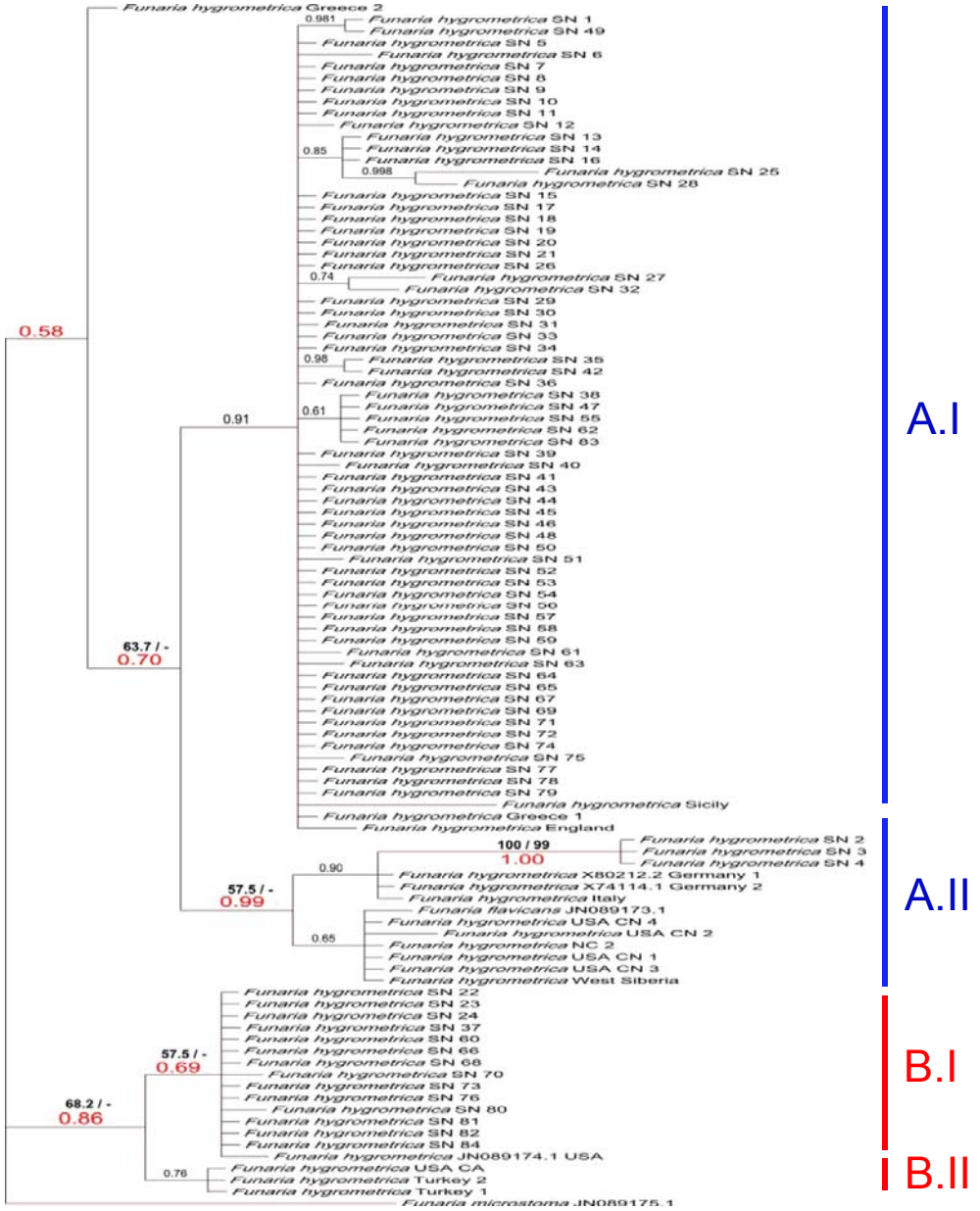

Figure 3.2 Bayesian inference rooted tree for ITS1 sequences based on BI method. Clades A, B and sub-clades are indicated and colored in accordance with the minimum spanning tree. NC stands for North Carolina, CA for California, CN for Connecticut (see Table 2.2, pp. 60), and SN for Sierra Nevada of Spain followed by sample number (see Table 3.1, pp. 119). For each clade BI probability values are given in red, while above 50 % bootstrap support values for MP and ML are superscripted in black, respectively.

Table 3.5 Genetic differentiation through AMOVA of *Funaria hygrometrica* based on ITS1 data. The source of variance (among and within locations), the degree of freedom (df), the sum of squares (SS), the variance components and the percentage of variation, are shown.

Source of variance	df	SS	Variance components	Percentage of variation
Among locations	14	293.67	2.98	37 %
Within locations	67	337.60	5.03	63 %
Total	81	631.28	8.01	

Φ_{ST} = 0.37 (p-value = 0.01)

3.1.2 NUCLEAR DNA – INTERNAL TRANSCRIBED SPACER 2 (ITS2)

3.1.2.1 Genetic variation and DNA polymorphism

The internal transcribed spacer 2 (ITS2) sequences for 84 *Funaria hygrometrica* samples from 17 locations ranged from 311 to 316 bp in size, separated into two groups of 67 occurrences with 316 bp and 17 occurrences with 311 to 312 bp in size (Table 3.6).

A total of seven haplotypes (GenBank accessions: JX985554 – JX985560) were defined by 56 variable sites with 8 mutations in 8 polymorphic (segregating) sites of which three were singletons (two-variants) and five were parsimony informative, while 48 gap-residues were inserted to keep the sequences aligned forming 21 indels.

The number of haplotypes *per* location ranged from one to five. Location 12 has the highest number of different haplotypes, followed by locations 7, 9 and 16 with two different haplotypes, while the rest of locations had only one haplotype sequence (Table 3.7).

When gaps were excluded, haplotype diversity was 0.37 and nucleotide diversity was 0.30 % for all samples combined together. Location 12 possessed the highest haplotype diversity (0.86) and nucleotide diversity of 0.50 %. Both locations 1 and 17 contained a single sample and therefore scored a haplotype diversity of 1.00 but with no nucleotide diversity. Locations 2, 3, 4, 5, 6, 8, 10, 11, 13, 14 and 15 had haplotype and nucleotide diversity values of 0.00 (Table 3.7).

A non-significant negative Tajima's D test value ($D = -1.13$, p-value > 0.10) confirmed the null hypothesis of neutral evolution and demographic equilibrium of the ITS2 spacer for all samples combined together. The neutral state of ITS2 was confirmed by Fu's test of a non-significant negative value ($Fs = -0.94$, p-value = 0.14).

Table 3.6 ITS2 data obtained from *Funaria hygrometrica*. Locations, Samples, haplotypes and size in bp of the ITS2 spacer, are shown.

Location	Sample	Haplotype	Size (bp)	Location	Sample	Haplotype	Size (bp)
1.1	1	1	316	10.1	43	1	316
2.1	2	2	311	10.2	44	1	316
2.2	3	2	311	10.3	45	1	316
2.3	4	2	311	10.4	46	1	316
3.1	5	1	316	10.5	47	1	316
3.2	6	1	316	10.6	48	1	316
4.1	7	1	316	10.7	49	1	316
4.2	8	1	316	10.8	50	1	316
4.3	9	1	316	10.9	51	1	316
5.1	10	1	316	10.10	52	1	316
5.2	11	1	316	11.1	53	1	316
5.3	12	1	316	11.2	54	1	316
5.4	13	1	316	11.3	55	1	316
5.5	14	1	316	11.4	56	1	316
5.6	15	1	316	11.5	57	1	316
5.7	16	1	316	11.6	58	1	316
5.8	17	1	316	11.7	59	1	316
6.1	18	1	316	11.8	60	1	316
6.2	19	1	316	11.9	61	1	316
6.3	20	1	316	11.10	62	1	316
6.4	21	1	316	12.1	63	5	312
7.1	22	3	312	12.2	64	1	316
7.2	23	4	312	12.3	65	1	316
7.3	24	3	312	12.4	66	3	312
8.1	25	1	316	12.5	67	1	316
8.2	26	1	316	12.6	68	3	312
8.3	27	1	316	12.7	69	6	311
8.4	28	1	316	12.8	70	7	316
8.5	29	1	316	13.1	71	1	316
8.6	30	1	316	13.2	72	1	316
8.7	31	1	316	13.3	73	1	316
8.8	32	1	316	13.4	74	1	316
8.9	33	1	316	13.5	75	1	316
9.1	34	1	316	13.6	76	1	316
9.2	35	1	316	14.1	77	2	311
9.3	36	1	316	14.2	78	2	311
9.4	37	3	312	15.1	79	2	311
9.5	38	1	316	15.2	80	2	311
9.6	39	1	316	16.1	81	1	316
9.7	40	1	316	16.2	82	3	312
9.8	41	1	316	16.3	83	1	316
9.9	42	1	316	17.1	84	3	312

Table 3.7 DNA polymorphism of ITS2 sequences. Sample locations, sample size (N), polymorphic *loci* (PL), number of haplotypes *per* location (h/loc), haplotype diversity (H_d), standard deviation of haplotype diversity (H_d s.d.), nucleotide diversity (π %) and standard deviation of nucleotide diversity (π % s.d.), are shown.

Location	N	PL*	h/loc	H_d	H_d s.d.	π %	π % s.d.
1	1	0	1	1.00	0.00	0.00	0.00
2	3	0	1	0.00	0.00	0.00	0.00
3	2	0	1	0.00	0.00	0.00	0.00
4	3	0	1	0.00	0.00	0.00	0.00
5	8	0	1	0.00	0.00	0.00	0.00
6	4	0	1	0.00	0.00	0.00	0.00
7	3	1	2	0.67	0.31	0.24	0.03
8	9	0	1	0.00	0.00	0.00	0.00
9	9	3	2	0.22	0.17	0.20	0.11
10	10	0	1	0.00	0.00	0.00	0.00
11	10	0	1	0.00	0.00	0.00	0.00
12	8	7	5	0.86	0.11	0.50	0.28
13	6	0	1	0.00	0.00	0.00	0.00
14	2	0	1	0.00	0.00	0.00	0.00
15	2	0	1	0.00	0.00	0.00	0.00
16	3	3	2	0.67	0.31	0.59	0.45
17	1	0	1	1.00	0.00	0.00	0.00
All samples	84	8	7	0.37*		0.30*	

***** Calculated when indels were excluded

Haplotype 1 was the most abundant (66 samples) and occupied a wide geographic distribution (12 locations), followed by haplotype 3 with seven samples and present in five locations. Haplotype 2, present in seven samples, was unique *per* locations 2, 14 and 15. The rest of haplotypes occurred only once and were restricted to one location. Only locations 9, 12 and 16 shared the two most abundant haplotypes (1 and 3) (Table 3.8).

Table 3.8 Frequencies of seven ITS2 haplotypes in 17 locations and the number of samples (S) counted for each haplotype. The frequency of each haplotype *per* location is equal to the number of samples of the same haplotype divided by the total number of samples found in that location.

	Haplotype						
Location	**1**	**2**	**3**	**4**	**5**	**6**	**7**
1	1.00	-	-	-	-	-	-
2	-	1.00	-	-	-	-	-
3	1.00	-	-	-	-	-	-
4	1.00	-	-	-	-	-	-
5	1.00	-	-	-	-	-	-
6	1.00	-	-	-	-	-	-
7	-	-	0.67	0.33	-	-	-
8	1.00	-	-	-	-	-	-
9	0.89	-	0.11	-	-	-	-
10	1.00	-	-	-	-	-	-
11	1.00	-	-	-	-	-	-
12	0.38	-	0.25	-	0.13	0.13	0.13
13	1.00	-	-	-	-	-	-
14	-	1.00	-	-	-	-	-
15	-	1.00	-	-	-	-	-
16	0.67	-	0.33	-	-	-	-
17	-	-	1.00	-	-	-	-
S	66	7	7	1	1	1	1

3.1.2.2 Population structure and phylogeny

3.1.2.2.1 Haplotype level

A minimum spanning tree (Fig. 3.3) showed a separation of the haplotypes into three groups. Group A, present in 67 samples, included haplotype 1 (the most abundant haplotype, present in 66 samples), together with haplotype 7, represented in blue color. Group B, present in 9 samples, included haplotype 3, present in 7 samples, together with haplotypes 4 and 5 (one sample each), and is marked in red color. Group C, present in 8 samples, included haplotype 2, present in 7 samples, together with haplotype 6, is separated by 12 mutational steps from group A (in which 8 indels of 12 bp in total were included) and 42 mutational steps from group B (in which 19 indels of 37 bp in total were included), and is marked in green. Small indels and divergences were present among and within each haplotype group.

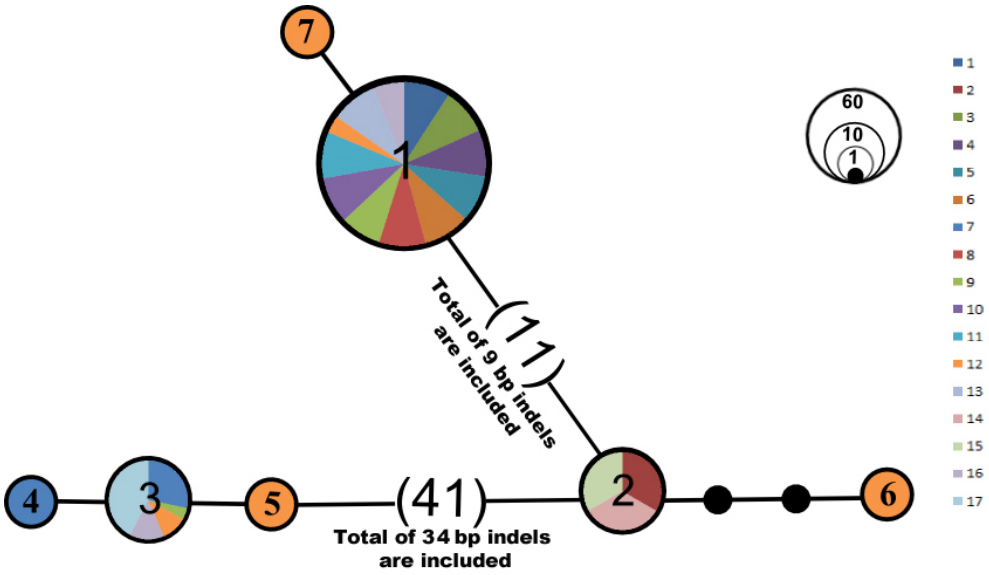

Figure 3.3 Minimum spanning tree of seven ITS2 haplotypes from 84 samples in 17 locations. The size of the circles is proportional to the haplotype abundance. Small black circles are hypothetical intermediate haplotypes not found in our study. Group A (blue), B (red) and C (green) are illustrated in the figure, while the number of hypothetical intermediate samples is shown between brackets with indication of the indels portion to the demonstrated distance.

For each haplotype, its secondary structure was folded directly without a template with an energy required for structure stability that ranged from -107.2 to -126.1. Folded structures were compared to the ITS2 database embeded models, resulting that gaps were repositioned according to the ITS2 Viridiplantae Hidden Markov Model (HMM) with the higest similarity score to the secondary structure template of *Funaria hygrometrica* (GenBank: X74114). The haplotype identity with respect to the proposed template ranged from 84.80 % to 97.80 %. The minimum identity (84.80 %) scored for haplotypes 3, 4 and 5. The maximum identities (97.80 and 96.00 %) were scored for haplotypes 2 and 6, respectively. In between, haplotypes 1 and 7 scored 93.90 and 93.30 %, respectively. According to the ITS database, the Viridiplantae HMM model is composed of a four helix structure (Fig. 3.4). The structure similarity was compared for each helix *per* haplotype. The helix similarity in descending order was the following, haplotypes 2, 6, 1, 7, 3, 5 and finally 4 (Table 3.9).

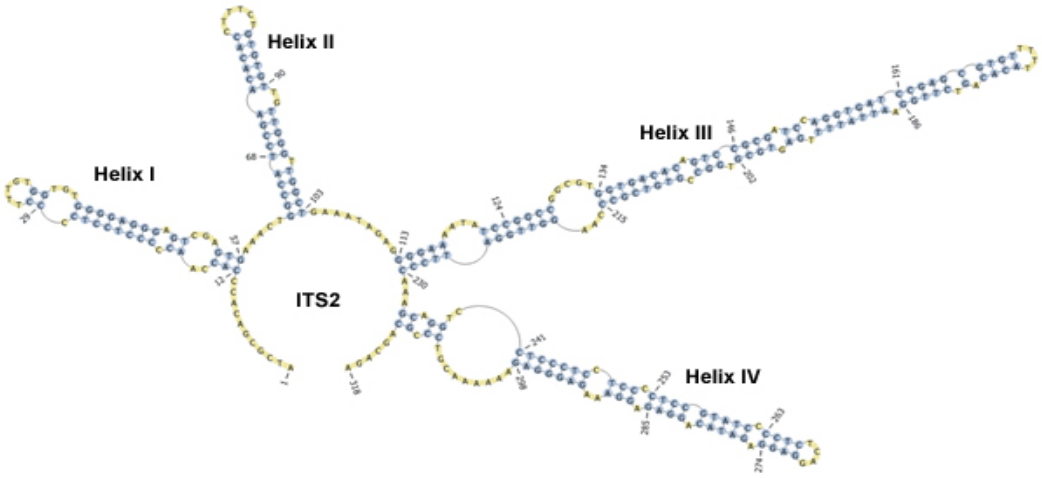

Figure 3.4 *Funaria hygrometrica* (GenBank: X74114) secondary structure template for ITS2 sequences. It is formed of four helical structures. Blue color is used for stems and yellow color regions for loops.

Table 3.9 Haplotype secondary structure data obtained from the ITS2 database. Structure stability energy (SE), percentage of haplotype identity with respect to the proposed template (IT %) and the helical transfer percentage for each helix, are shown.

Haplotype	SE	IT %	Helical transfer %			
			Helix 1	Helix 2	Helix 3	Helix 4
1	-125.5	93.90	100	100	93.47	96.29
2	-115.5	97.80	100	100	100	100
3	-107.3	84.80	100	81.25	91.30	74.07
4	-107.2	84.80	100	81.25	91.30	70.37
5	-107.2	84.80	100	81.25	91.30	74.07
6	-110.4	96.00	100	93.75	97.82	100
7	-126.1	93.30	100	100	93.47	96.29

A neighbour joining (NJ) tree was generated based on the secondary structure of each haplotype group (Fig. 3.5). The tree was rooted with *Funaria microstoma* (GenBank: JN089175.1). Two main clades could be clearly distinguished. Clade 1, included group B haplotypes as previously described by the minimum spanning tree (Fig. 3.3) and were geneticaly more derived than the clade 2. Clade 2 included both groups A and C, previously described by the minimum spanning tree. In contrast to the minimum spanning tree, in this analysis group C occupies a derived position and group A is more basal.

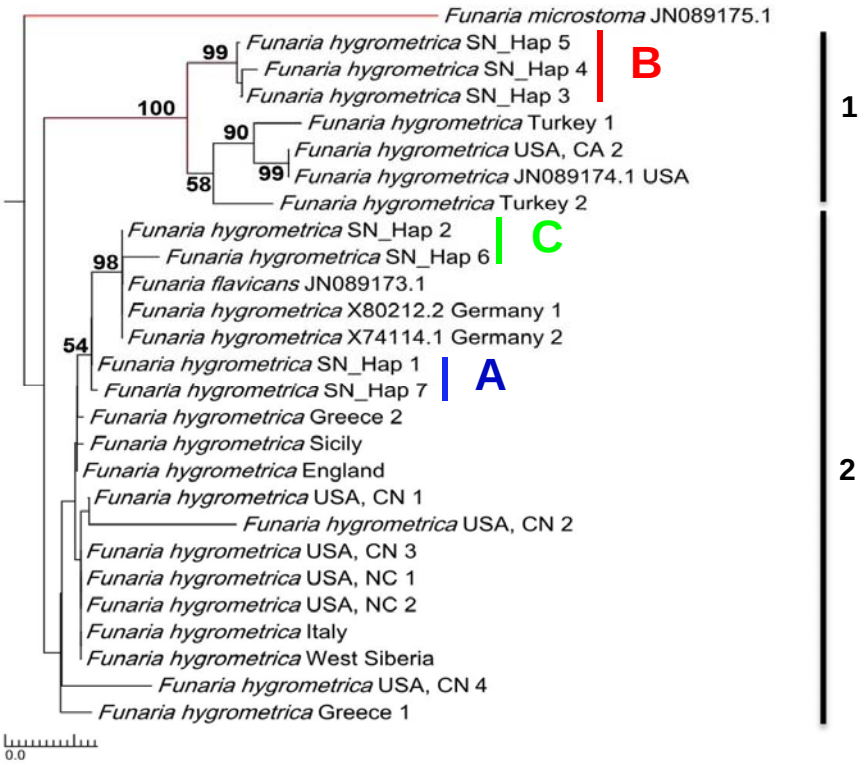

Figure 3.5 Rooted neighbor joining tree for ITS2 haplotypes based on secondary structure analysis. Groups A, B and C are indicated in accordance with the minimum spanning tree. NC stands for North Carolina, CA for California and CN for Connecticut (see Table 2.2, pp. 60), and SN for Sierra Nevada of Spain followed by the haplotype number (see Table 3.6, pp. 129). Bootstrap values above 50 % are indicated.

After excluding the gaps, a pairwise distance matrix between seven haplotypes was generated with MEGA based on the maximum composite likelihood model. The minimum genetic distance (0.32 %) was scored from the comparison between haplotypes 1 – 7 and 3 – (4 and 5). The maximum genetic distance value (3.15 %) was observed comparing haplotypes 4 – 6 (Table 3.10).

Table 3.10 Pairwise distance matrix between seven ITS2 haplotypes. Values are given in percentages, while minimum and maximum values are written in bold.

Haplotype	1	2	3	4	5	6	7
1							
2	0.33						
3	1.03	1.72					
4	1.38	2.08	**0.32**				
5	1.03	1.72	**0.32**	0.65			
6	1.32	0.97	2.79	**3.15**	2.79		
7	**0.32**	0.65	1.38	1.73	1.38	0.98	

3.1.2.2.2 Sample level

The phylogenetic tree of ITS2 sequences generated using BI was confirmed by both MP and ML methods (Fig. 3.6). The tree was rooted with *Funaria microstoma* (GenBank: JN089175.1). Two main clads could be clearly distinguished.

Clade A was supported with 0.99 of BI posterior probability, 79.7 bootstrap value of MP method and 92 bootstrap value of ML method. Sub-clade A.I was formed by Sierra Nevada samples of haplotypes belong to group C of the minimum spanning tree (Fig. 3.3) along with samples from Greece (sample number 2), U.S.A. (Connecticut: CN 1, 2, 3 and 4; North Carolina: NC 1 and 2), West Siberia, Italy, German accessions (GenBank: X80212.2 and X74114.1) and *Funaria flavicans* (GenBank: JN089173.1). The rest of clade A was formed by Sierra Nevada samples of haplotypes belonging to group A of the minimum spanning tree along with samples from Greece (sample number 1), Italy (Sicily) and England.

Clade B was supported with 1.00 of BI posterior probability and 100 bootstrap value of MP method, was formed by samples from Sierra Nevada of haplotypes belonging to group B of the minimum spanning tree (Fig. 3.3) along with U.S.A. samples (GenBank: JN089174.1 and California, CA), and Turkey samples (numbers 1 and 2).

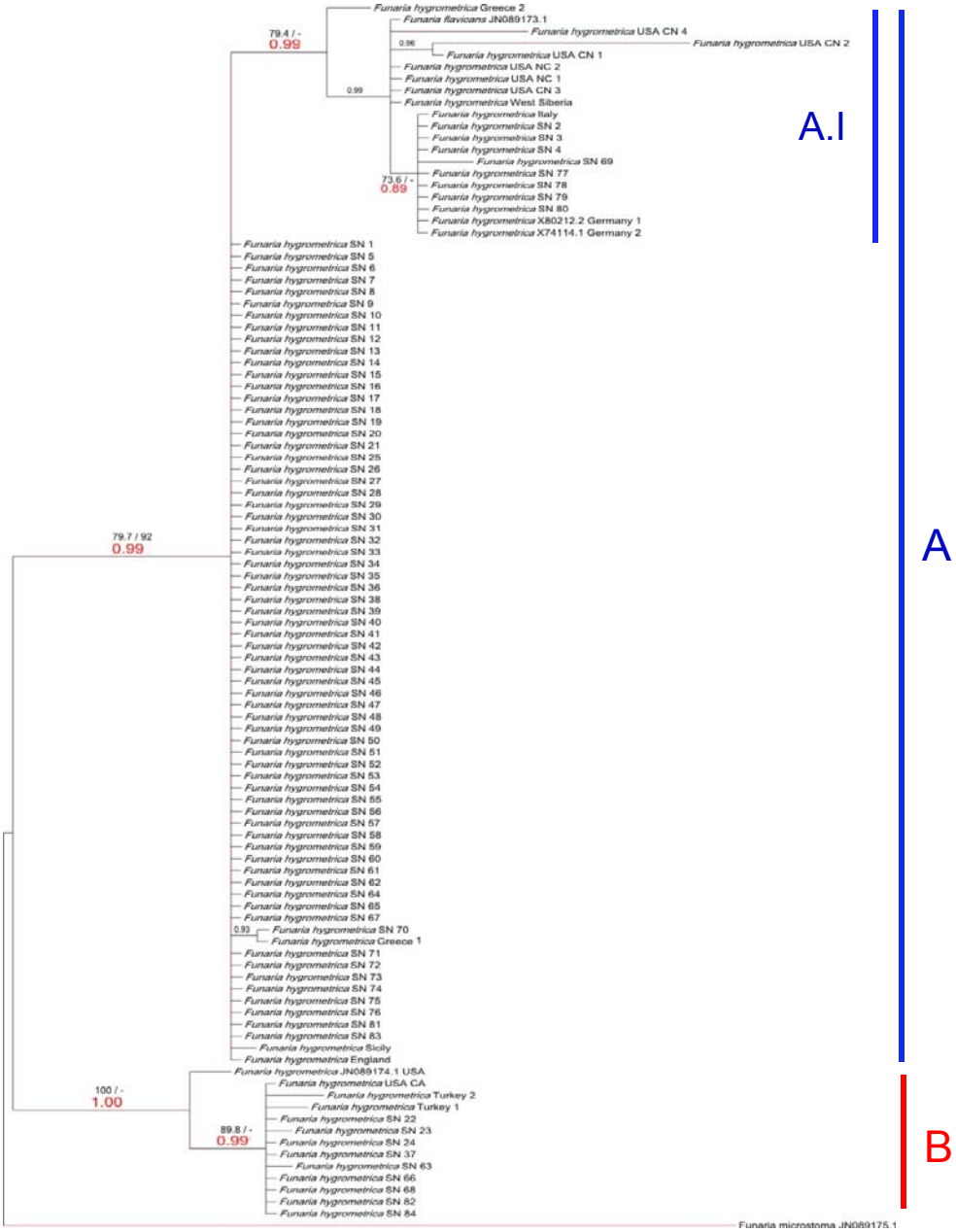

Figure 3.6 Rooted tree for ITS2 sequences based on BI method. Clades A and B are indicated and colored in accordance with the groups of the minimum spanning tree (sub-clade A.I represents group C). NC stands for North Carolina, CA for California, CN for Connecticut (see Table 2.2, pp. 60), and SN for Sierra Nevada of Spain followed by sample number (see Table 3.6, pp. 129). For each clade BI probability values are given in red, while above 50 % bootstrap support values for MP and ML are superscripted in black, respectively.

3.1.2.2.3 Location level

When locations 1 and 17 were excluded, a moderate gene flow (N_m) of 0.44 based on Phi-$_{ST}$ (Φ_{ST}) was calculated by GENALEX software.

An AMOVA test was estimated by GENALEX based on the pairwise differences, to test the genetic structure by comparing results from various groupings and choosing the structure that maximizes the among locations variance. Φ_{ST} was equal to 0.53, partitioned into a major genetic variation originated among locations, accounting for 53 % of the total variations, while 47 % of the genetic variation occurred within locations (Table 3.11).

Table 3.11 Genetic differentiation through AMOVA of *Funaria hygrometrica* based on ITS2 data. The source of variance (among and within locations), the degree of freedom (df), the sum of squares (SS), the variance components and the percentage of variation, are shown.

Source of variance	df	SS	Variance components	Percentage of variation
Among locations	14	241.38	2.77	53 %
Within locations	67	163.41	2.43	47 %
Total	81	404.80	5.20	

Φ_{ST} = 0.53 (p-value = 0.00)

3.1.3 CHLOROPLAST DNA

3.1.3.1 Genetic variation and DNA polymorphism

PCR amplification for the *rps*3-*rpl*16 region was successfully reached with high quality through the newly designed primers. The primers rps3-F1-Fun / rpl16-R3-Fun were finally used (see Table 2.3, pp. 68). Chloroplast *rps3-rpl16* region sequences for 79 *Funaria hygrometrica* samples from 17 locations ranged from 781 to 801 bp in size, separated into two groups of 12 occurrences with 781 to 782 bp and 67 occurrences with 797 to 801 bp in size (Table 3.12).

A total of 25 haplotypes (GenBank accessions: JX985507 – JX985531) were defined by 49 variable sites with 26 mutations in 25 polymorphic (segregating) sites of which 19 were singletons (18 of two-variants and one of three-variants) and 6 were parsimony informative, while 24 gap-residues were inserted to keep the sequences aligned forming 6 indels.

The number of haplotypes *per* location ranged from one to seven. Location 8 had the highest number of haplotypes, followed by location 10 with six haplotypes, and location 9 with four haplotypes. Locations 12 and 13 had three haplotypes. Locations 3, 4, 5, 6, 11, 14 and 17 had two haplotypes and the rest of locations had only one haplotype sequence (Table 3.13).

When gaps were excluded, haplotype diversity was 0.72 and nucleotide diversity was 0.21 % for all samples combined together. Locations 3, 8, 14 and 15 possessed the highest levels of haplotype diversity (1.00, 0.96, 1.00 and 1.00) and nucleotide diversity of 0.02, 0.03, 0.43 and 0.43 %, respectively. Both locations 1 and 17 contained a single sample, and therefore scored haplotype diversity of 1.00 but with no nucleotide diversity. Locations 2, 7 and 16 showed haplotype and nucleotide diversity value of 0.00 (Table 3.13).

Table 3.12 Chloroplast DNA data obtained from *Funaria hygrometrica*. Locations, samples, haplotypes and size in bp of *rps3-rpl16* region, are shown.

Location	Sample	Haplotype	Size (bp)	Location	Sample	Haplotype	Size (bp)
1.1	1	1	800	10.1	43	16	800
2.1	2	2	797	10.2	44	10	800
2.2	3	2	797	10.3	45	1	800
2.3	4	2	797	10.4	46	17	800
3.1	5	1	800	10.5	47	1	800
3.2	6	3	800	10.6	48	1	800
4.1	7	1	800	10.7	49	18	800
4.2	8	4	800	10.8	50	19	800
4.3	9	1	800	10.9	51	1	800
5.1	10	1	800	10.10	52	1	800
5.2	11	1	800	11.1	53	1	800
5.3	12	1	800	11.2	54	1	800
5.4	13	1	800	11.3	55	1	800
5.5	14	-	-	11.4	56	1	800
5.6	15	-	-	11.5	57	1	800
5.7	16	5	799	11.6	58	1	800
5.8	17	-	-	11.7	59	1	800
6.1	18	1	800	11.8	60	11	799
6.2	19	1	800	11.9	61	1	800
6.3	20	6	800	11.10	62	1	800
6.4	21	1	800	12.1	63	1	800
7.1	22	7	781	12.2	64	1	800
7.2	23	7	781	12.3	65	20	782
7.3	24	7	781	12.4	66	20	782
8.1	25	-	-	12.5	67	20	782
8.2	26	8	800	12.6	68	20	782
8.3	27	9	800	12.7	69	21	800
8.4	28	1	800	12.8	70	20	782
8.5	29	10	800	13.1	71	1	800
8.6	30	11	799	13.2	72	1	800
8.7	31	1	800	13.3	73	22	782
8.8	32	12	800	13.4	74	23	800
8.9	33	13	801	13.5	75	1	800
9.1	34	14	800	13.6	76	1	800
9.2	35	1	800	14.1	77	1	800
9.3	36	1	800	14.2	78	24	782
9.4	37	15	782	15.1	79	1	800
9.5	38	-	-	15.2	80	25	782
9.6	39	13	800	16.1	81	2	797
9.7	40	1	801	16.2	82	2	797
9.8	41	1	800	16.3	83	2	797
9.9	42	1	800	17.1	84	1	800

A significantly negative Tajima's D test value (D = -2.04, p-value < 0.05) rejected the null hypothesis of neutral evolution and/or demographic equilibrium of the cpDNA region. The un-neutral state of the cpDNA region was confirmed by Fu's test of a significant negative value (Fs = -15.40, p-value = 0.00).

Table 3.13 DNA polymorphism of cpDNA sequences. Sampling locations, sample size (N), polymorphic *loci* (PL), number of haplotypes *per* location (h/loc), haplotype diversity (H_d), standard deviation of haplotype diversity (H_d s.d.), nucleotide diversity (π %) and standard deviation of nucleotide diversity (π % s.d.), are shown.

Location	N	PL*	h/loc	H_d	H_d s.d.	π %	π % s.d.
1	1	0	1	1.00	0.00	0.00	0.00
2	3	0	1	0.00	0.00	0.00	0.00
3	2	1	2	1.00	0.50	0.02	0.02
4	3	3	2	0.66	0.31	0.04	0.03
5	5	0	2	0.40	0.23	0.01	0.01
6	4	1	2	0.50	0.26	0.01	0.01
7	3	0	1	0.00	0.00	0.00	0.00
8	8	5	7	0.96	0.07	0.04	0.02
9	8	6	4	0.64	0.18	0.12	0.07
10	10	6	6	0.77	0.13	0.02	0.02
11	10	0	2	0.20	0.15	0.05	0.01
12	8	2	3	0.61	0.16	0.22	0.12
13	6	5	3	0.60	0.21	0.16	0.10
14	2	3	2	1.00	0.50	0.43	0.44
15	2	0	2	1.00	0.50	0.43	0.44
16	3	0	1	0.00	0.00	0.00	0.00
17	1	3	1	1.00	0.00	0.00	0.00
All samples	79	25	25	0.72*		0.21*	

***** Calculated when indels were excluded

Haplotype 1 was the most abundant (41 samples) and occupied a wide geographic distribution (14 locations), followed by haplotype 2 with six samples and present in two locations and followed by haplotype 20 with five samples and present in only one location. Haplotype 7 with three samples occurred only in one location. Haplotypes 10, 11 and 13 occurred in two locations each. Haplotypes 3 to 9, 12 and 14 to 25 occurred only once. None of the locations shared the two most abundant haplotypes 1 and 2. Location 12 was the only one to share two haplotypes with high abundance (1 and 20) (Table 3.14).

3.1.3.2 Population structure and phylogeny

3.1.3.2.1 Haplotype level

A minimum spanning tree (Fig. 3.7) showed a separation of the haplotypes into three groups. Group A, present in 67 samples, included haplotype 1 (the most abundant haplotype, present in 41 samples), together with haplotypes 3 to 6, 8 to 14, 16 to 19, 21 and 23, and is represented in blue color. Group B, present in 12 samples, included haplotype 20 (the third most abundant haplotype, present in five samples), together with haplotypes 7, 15, 22, 24 and 25, and is marked in red color. The difference between the two groups was estimated as 19 mutational steps in which an indel of 17 bp was included. In between, group C, present in six samples, included only haplotype 2 (the second most abundant haplotype) at a distance from group A of six mutational steps in which an indel of 2 bp was included.

The three haplotype groups showed a clear divergence consisting in a loss of a fragment of 17 nucleotides. Small differences were present within each haplotype group, and corresponded clearly to the two size-ranges in base pairs (see Table 3.12). Secondary structure prediction showed that the fragment loss represents a partial stem. That loss might have ocurred due to one or more mutational events. Therefore, uncertainly, the 18 intermediate hypothetical samples in the haplotype minimum spanning tree might not exist (Fig. 3.8).

After excluding the gaps, a pairwise distance matrix between 25 haplotypes was generated with MEGA based on the maximum composite likelihood model. The minimum genetic distance (0.00 %) was scored from the comparison between haplotypes 1 – (5, 11 and 13), 5 – (11 and 13), 7 – 20 and 11 – 13. These haplotypes difer only in indels. The maximum genetic distances value (1.00 %) was observed comparing haplotypes 2 – 14 (Table 3.15).

Table 3.14 Frequencies of 25 cpDNA haplotypes in 17 locations and the number of samples (S) counted for each haplotype. The frequency of each haplotype *per* location is equal to the number of samples of the same haplotype divided by the total number of samples found in that location.

Haplotype	Location																	S
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	
1	1.00	-	0.50	0.67	0.80	0.75	-	0.22	0.61	0.50	0.90	0.24	0.66	0.50	0.50	-	1.00	41
2	-	1.00	-	-	-	-	-	-	-	-	-	-	-	-	-	1.00	-	6
3	-	-	0.50	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1
4	-	-	-	0.33	-	-	-	-	-	-	-	-	-	-	-	-	-	1
5	-	-	-	-	0.20	-	-	-	-	-	-	-	-	-	-	-	-	1
6	-	-	-	-	-	0.25	-	-	-	-	-	-	-	-	-	-	-	1
7	-	-	-	-	-	-	1.00	-	-	-	-	-	-	-	-	-	-	3
8	-	-	-	-	-	-	-	0.13	-	-	-	-	-	-	-	-	-	1
9	-	-	-	-	-	-	-	0.13	-	-	-	-	-	-	-	-	-	1
10	-	-	-	-	-	-	-	0.13	-	0.10	-	-	-	-	-	-	-	2
11	-	-	-	-	-	-	-	0.13	-	-	0.10	-	-	-	-	-	-	2
12	-	-	-	-	-	-	-	0.13	-	-	-	-	-	-	-	-	-	1
13	-	-	-	-	-	-	-	0.13	0.13	-	-	-	-	-	-	-	-	2
14	-	-	-	-	-	-	-	-	0.13	-	-	-	-	-	-	-	-	1
15	-	-	-	-	-	-	-	-	0.13	-	-	-	-	-	-	-	-	1
16	-	-	-	-	-	-	-	-	-	0.10	-	-	-	-	-	-	-	1
17	-	-	-	-	-	-	-	-	-	0.10	-	-	-	-	-	-	-	1
18	-	-	-	-	-	-	-	-	-	0.10	-	-	-	-	-	-	-	1
19	-	-	-	-	-	-	-	-	-	0.10	-	-	-	-	-	-	-	1
20	-	-	-	-	-	-	-	-	-	-	-	0.63	-	-	-	-	-	5
21	-	-	-	-	-	-	-	-	-	-	-	0.13	-	-	-	-	-	1
22	-	-	-	-	-	-	-	-	-	-	-	-	0.17	-	-	-	-	1
23	-	-	-	-	-	-	-	-	-	-	-	-	0.17	-	-	-	-	1
24	-	-	-	-	-	-	-	-	-	-	-	-	-	0.50	-	-	-	1
25	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.50	-	-	1

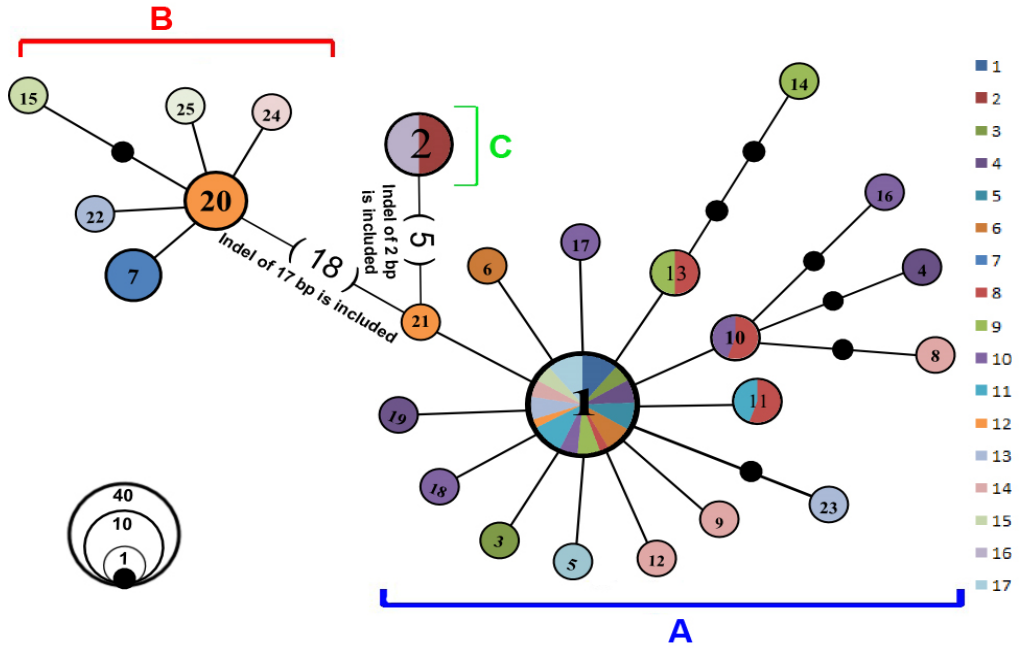

Figure 3.7 Minimum spanning tree of 25 cpDNA haplotypes from 79 samples in 17 locations. The size of the circles is proportional to the haplotype abundance. Small black circles represent intermediate haplotypes that are not in our study. Groups A (blue), B (red) and C (green) are illustrated in the figure, while the number of hypothetical intermediate mutations are shown between brackets with indication of the indels portion of the demonstrated distance.

3.1.3.2.2 Sample level

The phylogenetic tree of cpDNA sequences generated by BI was confirmed by both MP and ML methods (Fig. 3.9). An appropriate out-group was not found. Two main clades could be distinguished. Clade A was supported with 0.78 of BI posterior probability, 54 bootstrap value of MP method and 50 bootstrap value of ML method. It is formed by samples of haplotypes belong to group A, of the minimum spanning tree (Fig. 3.7). Clade B was supported with 0.60 of BI posterior probability and 58 bootstrap value of MP method. It was formed by samples of haplotypes belonging to groups B and C (samples 2, 3 and 4) of the minimum spanning tree. In between sample 37 (haplotype 15) was not included in any of the clades, however, it belongs to group B of the minimum spanning tree.

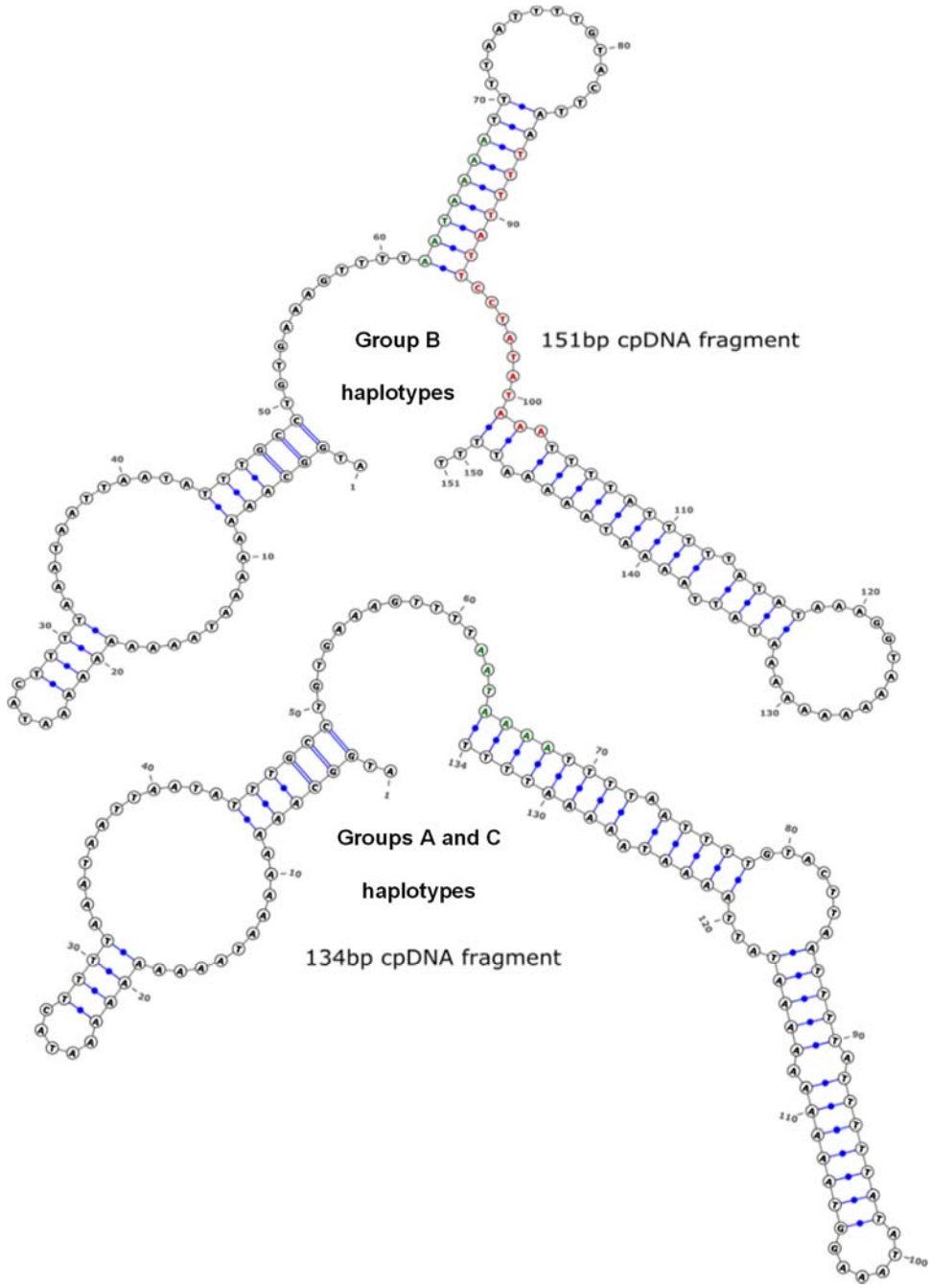

Figure 3.8 Secondary structure analysis of cpDNA sequences. Only 151 bp of the complete sequence are illustrated, in which the lost fragment is included. 17 nucleotides are labeled in red indicate the fragment present only in group B haplotypes, while 7 nucleotides labeled in green show the position of where the new structure started in groups A and C haplotypes.

Table 3.15 Pairwise distance matrix between 25 cpDNA haplotypes (H). Values are given in percentages, while minimum and maximum values are written in bold.

H	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1																									
2	0.50																								
3	0.13	0.63																							
4	0.38	0.89	0.50																						
5	**0.00**	0.50	0.13	0.38																					
6	0.13	0.63	0.25	0.50	0.13																				
7	0.26	0.26	0.39	0.64	0.26	0.39																			
8	0.38	0.89	0.50	0.50	0.38	0.50	0.64																		
9	0.13	0.63	0.25	0.50	0.13	0.25	0.39	0.50																	
10	0.13	0.63	0.25	0.25	0.13	0.25	0.39	0.25	0.25																
11	**0.00**	0.50	0.13	0.38	**0.00**	0.13	0.26	0.38	0.13	0.13															
12	0.13	0.63	0.25	0.50	0.13	0.25	0.38	0.50	0.25	0.25	0.13														
13	**0.00**	0.50	0.13	0.38	**0.00**	0.13	0.26	0.38	0.13	0.13	**0.00**	0.13													
14	0.50	**1.00**	0.63	0.88	0.50	0.38	0.77	0.88	0.63	0.63	0.50	0.63	0.50												
15	0.26	0.51	0.39	0.64	0.26	0.39	0.26	0.64	0.39	0.39	0.26	0.38	0.26	0.77											
16	0.38	0.89	0.50	0.50	0.38	0.50	0.64	0.50	0.50	0.25	0.38	0.50	0.38	0.88	0.64										
17	0.13	0.63	0.25	0.50	0.13	0.25	0.39	0.50	0.25	0.25	0.13	0.25	0.13	0.50	0.39	0.50									
18	0.13	0.63	0.25	0.50	0.13	0.25	0.39	0.50	0.25	0.25	0.13	0.25	0.13	0.63	0.39	0.50	0.25								
19	0.13	0.63	0.25	0.50	0.13	0.25	0.39	0.50	0.25	0.25	0.13	0.25	0.13	0.63	0.39	0.50	0.25	0.25							
20	0.26	0.26	0.38	0.64	0.26	0.38	**0.00**	0.64	0.38	0.38	0.26	0.38	0.26	0.77	0.26	0.64	0.38	0.38	0.38						
21	0.13	0.38	0.25	0.50	0.13	0.25	0.13	0.50	0.25	0.25	0.13	0.25	0.13	0.63	0.13	0.50	0.25	0.25	0.25	0.13					
22	0.38	0.39	0.51	0.51	0.39	0.51	0.13	0.51	0.51	0.26	0.39	0.51	0.38	0.90	0.38	0.51	0.51	0.51	0.51	0.13	0.26				
23	0.25	0.76	0.38	0.63	0.25	0.38	0.51	0.63	0.38	0.38	0.25	0.38	0.25	0.75	0.51	0.63	0.38	0.38	0.38	0.51	0.38	0.64			
24	0.38	0.39	0.51	0.77	0.39	0.51	0.13	0.77	0.51	0.51	0.39	0.51	0.38	0.90	0.38	0.77	0.51	0.51	0.51	0.13	0.26	0.26	0.64		
25	0.38	0.39	0.51	0.77	0.39	0.51	0.13	0.77	0.51	0.51	0.39	0.51	0.38	0.90	0.38	0.77	0.51	0.51	0.51	0.13	0.26	0.26	0.64	0.26	

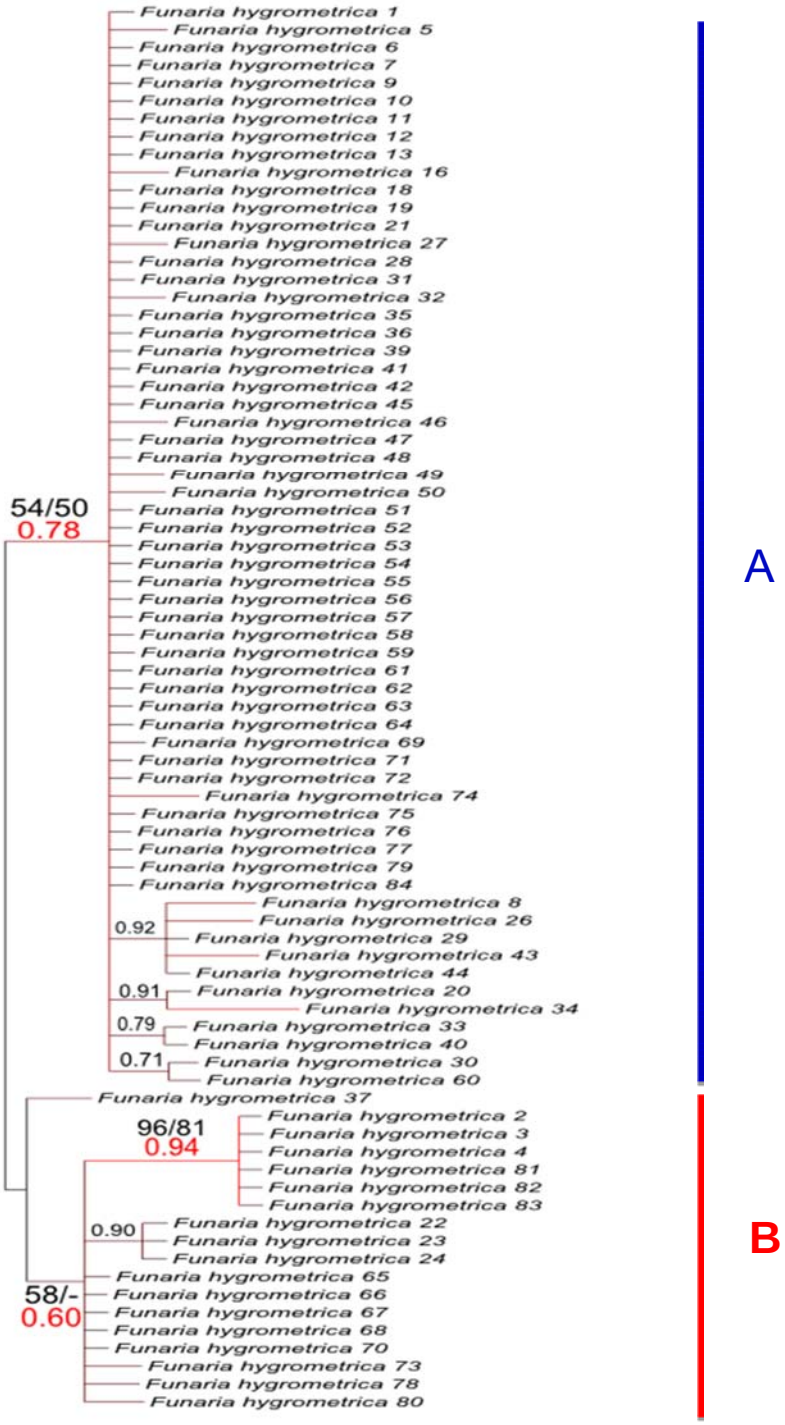

Figure 3.9 Unrooted tree for cpDNA sequences based on BI method. Clades A and B are indicated and colored in accordance with the groups of the minimum spanning tree. *Funaria hygrometrica* samples from Sierra Nevada are followed by sample number (see Table 3.12, pp. 139). For each clade, BI probability values are given in red, while above 50 % bootstrap support values for MP and ML are superscripted in black, respectively.

3.1.3.2.3 Location level

When locations 1 and 17 were excluded, a moderate gene flow (N_m) of 0.52 based on Phi-$_{ST}$ (Φ_{ST}) was calculated by GENALEX software.

An AMOVA test was estimated by GENALEX based on the pairwise differences, to test the genetic structure by comparing results from various groupings and choosing the structure that maximizes the among locations variance. Φ_{ST} was equal to 0.49, partitioned into a major genetic variation originated within locations, accounting for 51 % of the total variations, and was approximatly equal to the genetic variation occurring among locations (49 %; Table 3.16).

Table 3.16 Genetic differentiation through AMOVA of *Funaria hygrometrica* based on cpDNA data. The source of variance (among and within locations), the degree of freedom (df), the sum of squares (SS), the variance components and the percentage of variation, are shown.

Source of variance	df	SS	Variance components	Percentage of variation
Among locations	14	153.69	1.81	49 %
Within locations	62	116.96	1.88	51 %
Total	76	270.66	3.69	

Φ_{ST} = 0.49 (p-value = 0.00)

3.1.4 MITOCHONDRIAL DNA

3.1.4.1 Genetic variation and DNA polymorphism

Mitochondrial *rpl*5-*rpl*16 intergenic spacer sequences for 84 *Funaria hygrometrica* samples from 17 locations ranged from 790 to 836 bp in size, separated into two groups of 66 occurrences with 790 to 791 bp and 18 occurrences with 835 to 836 bp in size (Table 3.17).

A total of 11 haplotypes (GenBank accessions: JX912563 – JX912573) were defined by 60 variable sites with 12 mutations in 12 polymorphic (segregating) sites of which 11 were singletons and one was parsimony informative, while 48 gap-residues were inserted to keep the sequences aligned forming four indels.

The number of haplotypes *per* location ranged from one to four. Location 12 had the highest number of haplotypes, followed by location 13 with three haplotypes. Locations 4, 9, 10, 14, 15, and 16 had two haplotypes. The rest of locations had only one haplotype sequence (Table 3.18).

When gaps were excluded, haplotype diversity was 0.43 and nucleotide diversity was 0.03 % for all samples. Locations 14 and 15 possessed the highest haplotype diversity (1.00) and nucleotide diversity of 0.78 and 0.08 %, respectively. Both locations 1 and 17 contained a single sample, and therefore scored a haplotype diversity of 1.00 but with no nucleotide diversity. Locations 2, 3, 5, 6, 7, 8 and 11 showed a haplotype and nucleotide diversity of 0.00 (Table 3.18).

A significantly negative Tajima's D test value ($D = -2.37$, p-value < 0.01) rejected the null hypothesis of neutral evolution and/or demographic equilibrium of the mtDNA region. The un-neutral state of the cpDNA region was confirmed by Fu's test of a significant negative value ($Fs = -10.03$, p-value = 0.00).

Table 3.17 Mitochondrial DNA data obtained from *Funaria hygrometrica*. Locations, samples, haplotypes and size in bp of *rps5-rpl16* intergenic spacer, are shown.

Location	Sample	Haplotype	Size (bp)	Location	Sample	Haplotype	Size (bp)
1.1	1	1	791	10.1	43	1	791
2.1	2	2	836	10.2	44	1	791
2.2	3	2	836	10.3	45	1	791
2.3	4	2	836	10.4	46	1	791
3.1	5	1	791	10.5	47	1	791
3.2	6	1	791	10.6	48	1	791
4.1	7	1	791	10.7	49	1	791
4.2	8	1	791	10.8	50	1	791
4.3	9	3	790	10.9	51	1	791
5.1	10	1	791	10.10	52	7	791
5.2	11	1	791	11.1	53	1	791
5.3	12	1	791	11.2	54	1	791
5.4	13	1	791	11.3	55	1	791
5.5	14	1	791	11.4	56	1	791
5.6	15	1	791	11.5	57	1	791
5.7	16	1	791	11.6	58	1	791
5.8	17	1	791	11.7	59	1	791
6.1	18	1	791	11.8	60	1	791
6.2	19	1	791	11.9	61	1	791
6.3	20	1	791	11.10	62	1	791
6.4	21	1	791	12.1	63	1	791
7.1	22	2	836	12.2	64	1	791
7.2	23	2	836	12.3	65	2	836
7.3	24	2	836	12.4	66	5	836
8.1	25	1	791	12.5	67	6	836
8.2	26	1	791	12.6	68	2	836
8.3	27	1	791	12.7	69	1	791
8.4	28	1	791	12.8	70	2	836
8.5	29	1	791	13.1	71	1	791
8.6	30	1	791	13.2	72	4	791
8.7	31	1	791	13.3	73	1	791
8.8	32	1	791	13.4	74	1	791
9.1	33	1	791	13.5	75	8	790
9.2	34	1	791	13.6	76	1	791
9.3	35	1	791	14.1	77	1	791
9.4	36	1	791	14.2	78	9	836
9.5	37	2	836	15.1	79	10	836
9.6	38	1	791	15.2	80	11	835
9.7	39	1	791	16.1	81	2	836
9.8	40	1	791	16.2	82	2	836
9.9	41	1	791	16.3	83	1	791
9.10	42	1	791	17.1	84	2	836

Table 3.18 DNA polymorphism of mtDNA sequences. Sample locations, sample size (N), polymorphic *loci* (PL), number of haplotypes *per* location (h/loc), haplotype diversity (H_d), standard deviation of haplotype diversity (H_d s.d.), nucleotide diversity (π %) and standard deviation of nucleotide diversity (π % s.d.), are shown.

Location	N	PL*	h/loc	H_d	H_d s.d.	π %	π % s.d.
1	1	0	1	1.00	0.00	0.00	0.00
2	3	0	1	0.00	0.00	0.00	0.00
3	2	0	1	0.00	0.00	0.00	0.00
4	3	1	2	0.66	0.31	0.08	0.09
5	8	0	1	0.00	0.00	0.00	0.00
6	4	0	1	0.00	0.00	0.00	0.00
7	3	0	1	0.00	0.00	0.00	0.00
8	9	0	1	0.00	0.00	0.00	0.00
9	9	0	2	0.22	0.16	0.16	0.09
10	10	1	2	0.20	0.15	0.01	0.02
11	10	0	1	0.00	0.00	0.00	0.00
12	8	4	4	0.78	0.11	0.41	0.23
13	6	1	3	0.60	0.21	0.04	0.04
14	2	2	2	1.00	0.50	0.78	0.79
15	2	4	2	1.00	0.50	0.08	0.09
16	3	0	2	0.66	0.31	0.50	0.38
17	1	0	1	1.00	0.00	0.00	0.00
All samples	84	12	11	0.43*		0.03*	

***** Calculated when indels were excluded

Haplotype 1 was the most abundant (62 samples) and occupied a wide geographic distribution (13 locations), followed by haplotype 2 with 13 samples and present in six locations. Haplotypes 3 to 11 only occurred once and were restricted to one location. Only locations 9, 12, and 16 shared the two most abundant haplotypes (1 and 2), particularly location 12 had two more haplotypes (5 and 6) (Table 3.19).

Table 3.19 Frequencies of 11 mtDNA haplotypes in 17 locations and the number of samples (S) counted for each haplotype. The frequency of each haplotype *per* location is equal to the number of samples of the same haplotype divided by the total number of samples found in that location.

	Haplotype										
Location	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
1	1.00	-	-	-	-	-	-	-	-	-	-
2	-	1.00	-	-	-	-	-	-	-	-	-
3	1.00	-	-	-	-	-	-	-	-	-	-
4	0.67	-	0.33	-	-	-	-	-	-	-	-
5	1.00	-	-	-	-	-	-	-	-	-	-
6	1.00	-	-	-	-	-	-	-	-	-	-
7	-	1.00	-	-	-	-	-	-	-	-	-
8	1.00	-	-	-	-	-	-	-	-	-	-
9	0.89	0.11	-	-	-	-	-	-	-	-	-
10	0.90	-	-	0.10	-	-	-	-	-	-	-
11	1.00	-	-	-	-	-	-	-	-	-	-
12	0.38	0.38	-	-	0.12	0.12	-	-	-	-	-
13	0.68	-	-	-	-	-	0.16	0.16	-	-	-
14	0.50	-	-	-	-	-	-	-	0.50	-	-
15	-	-	-	-	-	-	-	-	-	0.50	0.50
16	0.37	0.66	-	-	-	-	-	-	-	-	-
17	-	1.00	-	-	-	-	-	-	-	-	-
S	62	13	1	1	1	1	1	1	1	1	1

3.1.4.2 Population structure and phylogeny

3.1.4.2.1 Haplotype level

A minimum spanning tree (Fig. 3.10) showed a separation of the haplotypes into two main groups. Group A, present in 66 samples, included haplotype 1 (the most abundant haplotype, present in 62 samples), together with haplotypes 3, 4, 7, and 8 (one sample each), and marked in blue color. Group B, present in 18 samples, included haplotype 2 (the second most abundant haplotype, present in 13 samples), together with haplotypes 5, 6, 9, 10, and 11 (one sample each), and marked in red color. The difference between the two groups was estimated as 45 mutational steps, due to an indel of 45 bp.

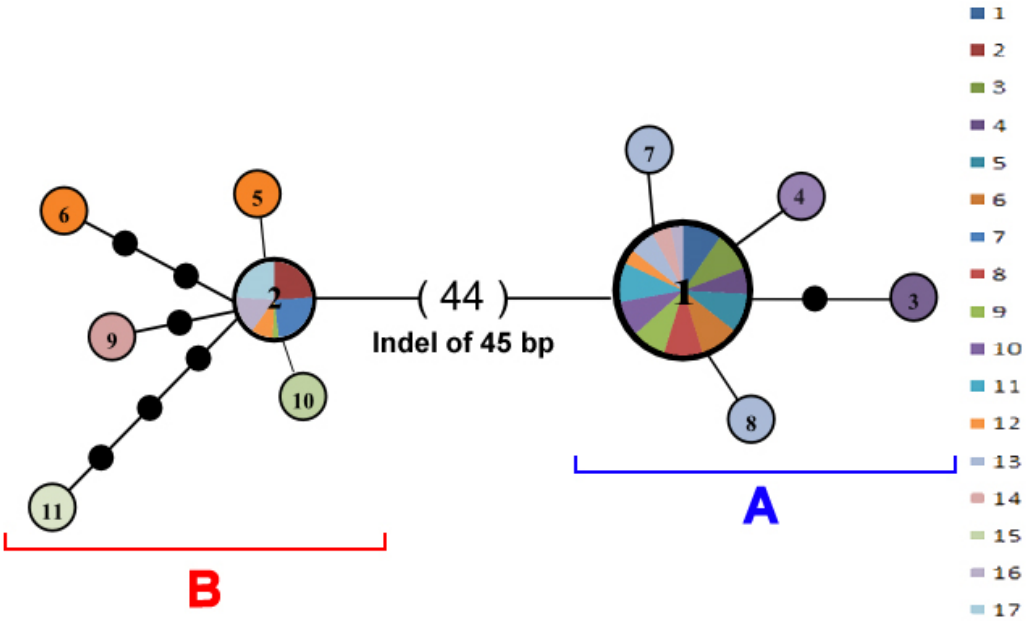

Figure 3.10 Minimum spanning tree of 11 mtDNA haplotypes of 84 samples in 17 locations. The size of the circles is proportional to the haplotype abundance. Small black circles are hypothetical intermediate haplotypes not found in our study. Groups A (blue) and B (red) are illustrated in the figure, while the number of hypothetical intermediate samples are shown between brackets with indication of the indels portion to the demonstrated distance.

The two haplotype groups were present in all locations. The major difference consisted in a fragment loss of 45 nucleotides, while small differences were present in each haplotype group. A secondary structure prediction showed that the fragment loss represented a complete loop, which probably indicates the loss as a single event. Therefore, the 44 intermediate hypothetical samples in the haplotype minimum spanning tree do not really exist (Fig. 3.11).

After excluding the gaps, a pairwise distance matrix between the 11 haplotypes was generated with MEGA based on the maximum composite likelihood model. The minimum genetic distances (0.00 %) was scored from the comparison between haplotypes 1 – (2 and 8), and 2 – 8, which difer only in indels. The maximum genetic distance value (0.72 %) was observed comparing haplotypes 6 – 11 (Table 3.20).

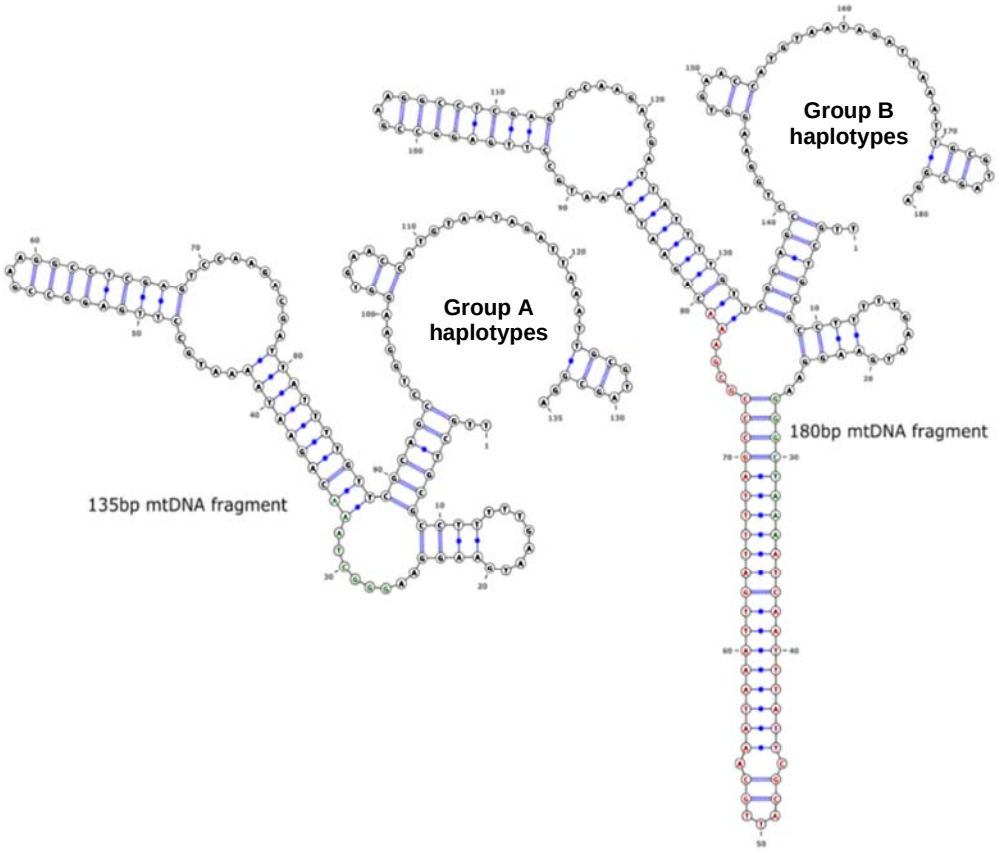

Figure 3.11 Secondary structure analysis of mtDNA sequences. Secondary structure predicated for the two groups of haplotypes according to the minimum spanning tree. Only 180 bp of the complete sequence are illustrated in which the lost fragment is included. The 45 nucleotides corresponding to the large indel are labeled in red to indicate the fragment present in group B, while seven nucleotides labeled in green show the new position taken to recover the gap after the loop was lost.

3.1.4.2.2 Sample level

The phylogenetic tree of mtDNA sequences generated by BI was confirmed by both MP and ML methods (Fig. 3.12). Two clades were observed, which corresspond to the two groups illustrated by the minimum spanning tree (Fig. 3.10), as grouped samples were of the same haplotype group.

Table 3.20 Pairwise distance matrix between 11 mtDNA haplotypes (H). Values are given in percentages, while minimum and maximum values are written in bold.

H	1	2	3	4	5	6	7	8	9	10	11
1											
2	**0.00**										
3	0.13	0.13									
4	0.13	0.13	0.25								
5	0.13	0.12	0.25	0.25							
6	0.38	0.36	0.51	0.51	0.48						
7	0.13	0.13	0.25	0.25	0.25	0.51					
8	**0.00**	**0.00**	0.13	0.13	0.13	0.38	0.13				
9	0.25	0.24	0.38	0.38	0.36	0.36	0.38	0.25			
10	0.13	0.12	0.25	0.25	0.24	0.48	0.25	0.13	0.36		
11	0.38	0.36	0.51	0.51	0.48	**0.72**	0.51	0.38	0.60	0.48	

Clade A was supported with 0.68 of BI posterior probability, 58 and 68 bootstrap value of MP and ML, respectively. It included samples with the most abundant haplotypes (blue).

Clade B was supported with 1.00 of BI posterior probability, 99 and 99 bootstrap value of MP and ML, respectively. It included samples with the less abundant haplotypes (red) along with the *Funaria hygrometrica* sample obtained from GenBank (JN584637.1).

3.1.4.2.3 Location level

When locations 1 and 17 were excluded, a low gene flow (N_m) of 0.27 based on Phi-$_{ST}$ (Φ_{ST}) was calculated by GENALEX software.

An AMOVA test was estimated by GENALEX based on the pairwise differences, to test the genetic structure by comparing results from various groupings and choosing the structure that maximizes the among locations variance. Φ_{ST} was equal to 0.65, partitioned into a major genetic variation originated among locations, accounting for 65 % of the total variations, while 35 % of the genetic variation occurred within locations (Table 3.21).

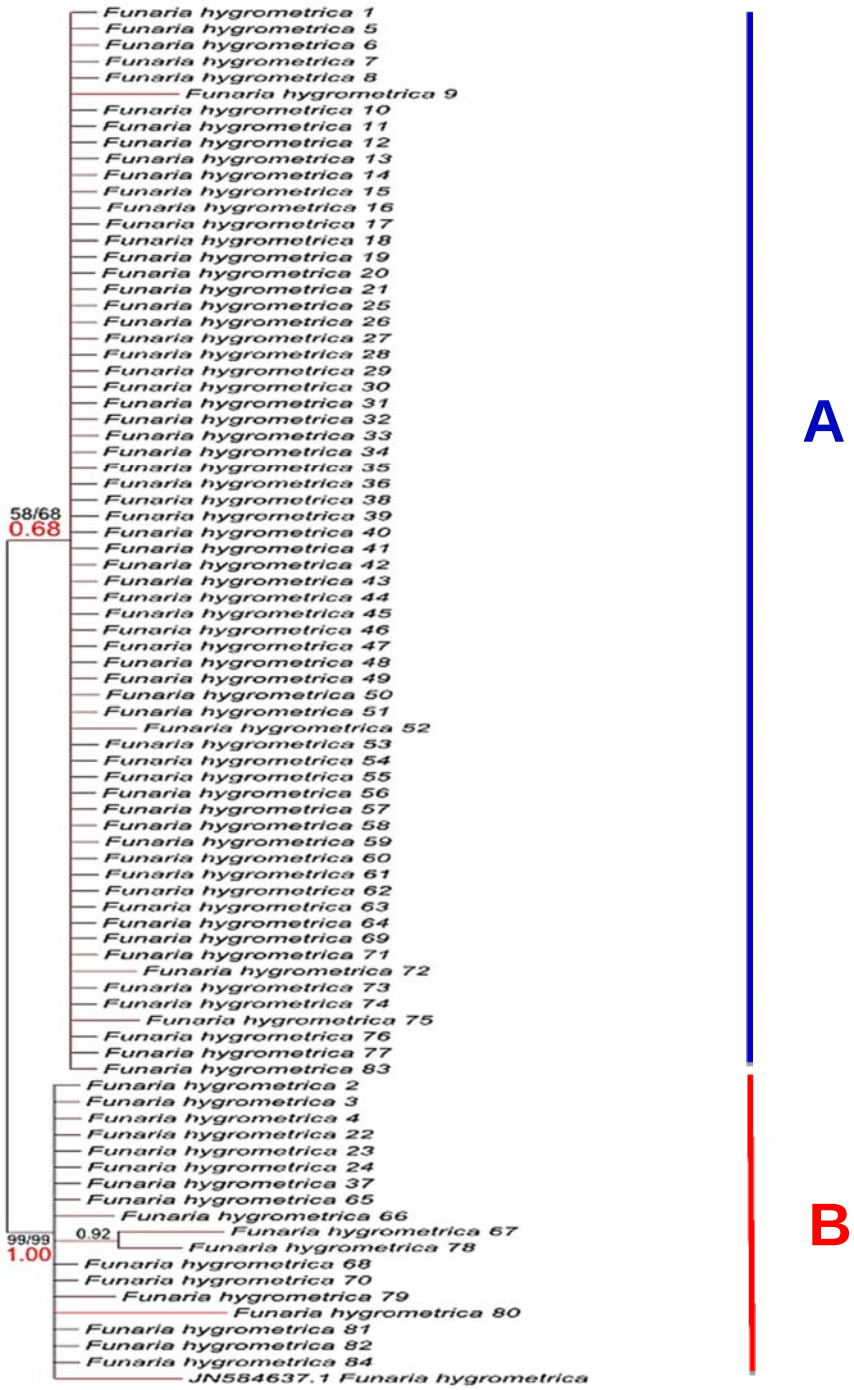

Figure 3.12 Unrooted tree for mtDNA sequences based on BI method. Clades A and B are indicated and colored in accordance with the minimum spanning tree. *Funaria hygrometrica* from Sierra Nevada followed by sample number (see Table 3.17, pp. 149). For each clade BI probability values are given in red, while above 50 % bootstrap support values for MP and ML are superscripted in black, respectively.

Table 3.21 Genetic differentiation through AMOVA of *Funaria hygrometrica* based on mtDNA data. The source of variance (among and within locations), the degree of freedom (df), the sum of squares (SS), the variance components and the percentage of variation, are shown.

Source of variance	df	SS	Variance components	Percentage of variation
Among locations	14	433.92	5.27	65 %
Within locations	67	187.27	2.79	35 %
Total	81	621.19	8.07	

Φ_{ST} = 0.65 (p-value = 0.00)

3.1.5 MICROSATELLITE

3.1.5.1 Amplification, scoring and readability

PCR amplification was successful for five SSR regions. High quality was successfully reached for all samples (e.g. sample 4; Fig. 3.13 A, B). All peak analyses and automated band scoring were successful. More than 1 fragment *per locus* for different samples were detected (e.g. green peaks; Fig. 3.13 A) within the reference range (see Table 2.4, pp. 72).

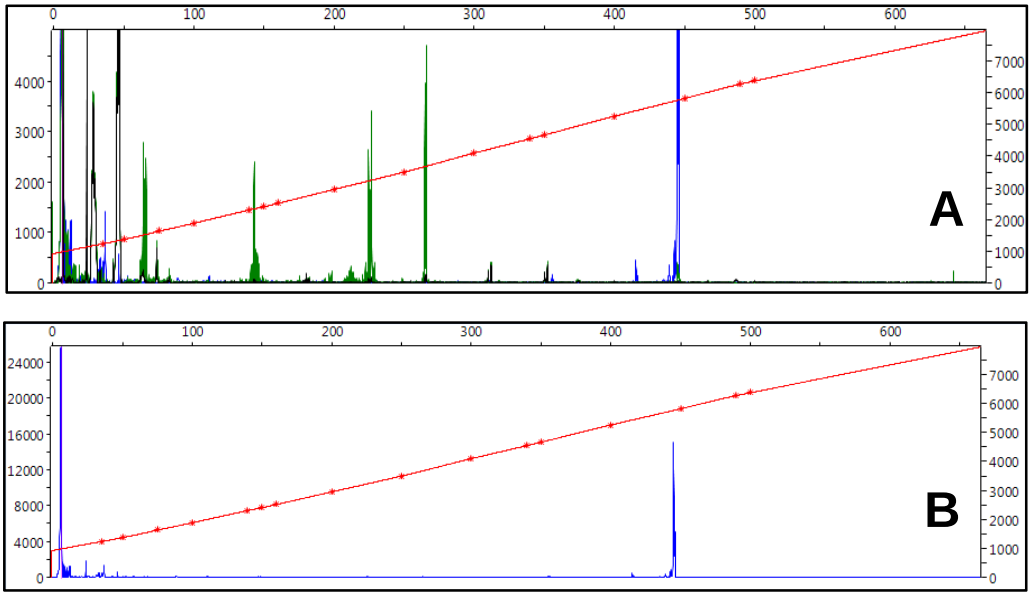

Figure 3.13 PEAKSCANNER chromatogram of amplified microsatellite *loci* for sample 4; A: peaks of M2 (blue), M3 (green) and M4 (black); B: peaks of the M2 *locus* (blue).

3.1.5.2 Population data, total and within locations genetic variation

The genetic diversity was calculated starting by the first scored allele and increasing by one until all scored alleles were included. The results showed the ability of the scored bands to calculate efficiently the genetic diversity for all samples combined together. Plotting SSR *loci* against their calculated genetic diversity, the curve approximately stopped at maximum mean of genetic diversity (1.00), however, it did not reach an obvious plateau. This indicates that the current scored bands were almost sufficient to estimate the genetic diversity (Fig. 3.14).

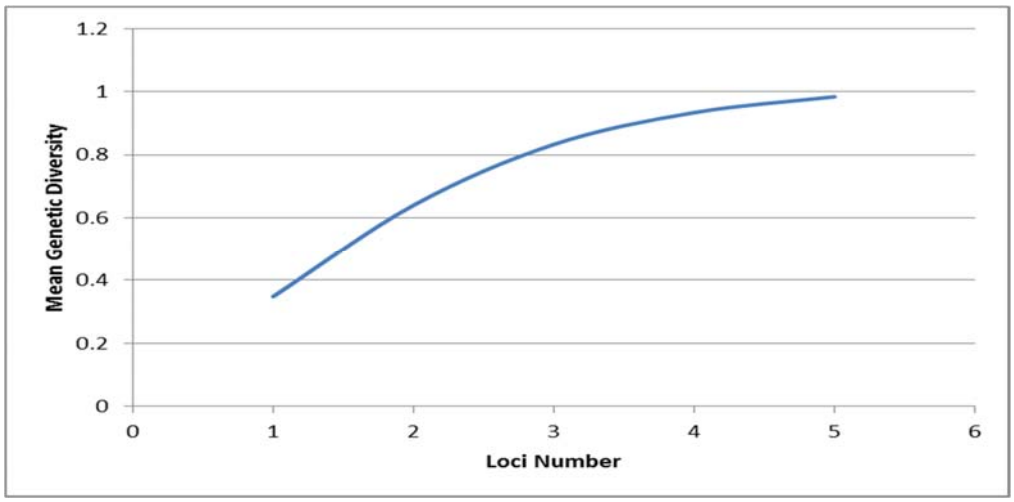

Figure 3.14 Five SSR *loci* are plotted against mean genetic diversity calculated for 84 samples in 17 locations.

A total of 19 alleles were scored from the five SSR regions for all the 84 samples. All scored alleles were polymorphic. The unbiased gene diversity (UH_z) for all samples combined together was 0.45. The maximum and minimum UH_z were 0.60 for location 16 and 0.23 for location 7, respectively. The observed heterozygosity (OH_z) for all samples combined together was 0.53. The maximum and minimum OH_z were 0.63 for location 12 and 0.13 for location 2, respectively. The number of alleles *per locus* (n_A) for all samples combined together was 3.8. The maximum and minimum n_A were 3.20 for location 5 and 1.40 for location 7, respectively. The effective number of alleles (n_e) for all samples combined together was 2.75. The maximum and minimum n_e were 2.58 for location 5 and 1.36 for location 7, respectively. The value of polymorphic index content (PIC) for all samples combined together was 0.41. The maximum and minimum PIC were 0.42 for location 16 and 0.14 for location 7, respectively (Table 3.22). The locations with maximum and minimum values of n_A, n_e and UH_z were location 5 and location 7, respectively (Fig. 3.15). The number of scored alleles *per* location ranged from 7 to 15. Locations 5 and 9 had the highest number of different alleles, while locations 1 and 7 had the lowest number of alleles (Table 3.23). Each location showed a unique allelic frequencies pattern for the studied SSR loci (Fig. 3.16).

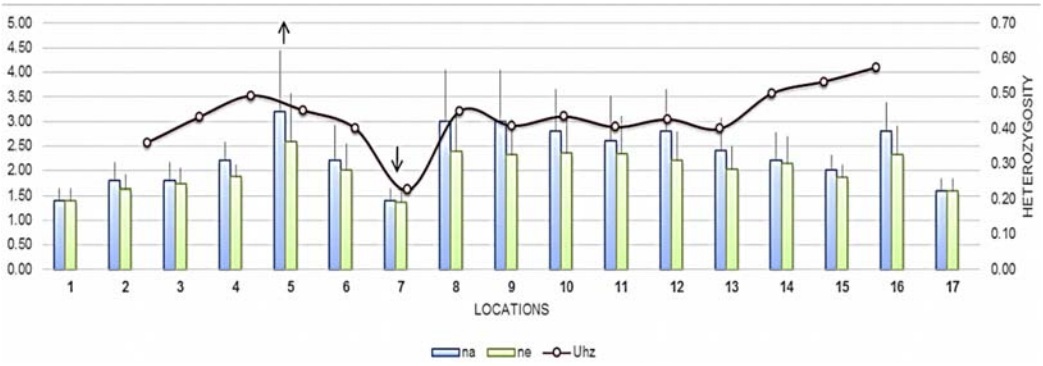

Figure 3.15 Allelic patterns across locations. The number of alleles per *locus* (n_A), number of effective alleles (n_e) and unbiased gene diversity (UH_z) for each location, are shown. Locations with maximum and minimum values are indicated with upwards and downwards arrow, respectively.

Table 3.22 Genetic variation of SSR molecular marker *per* location. Sample locations, sample size (N), unbiased gene diversity (UH_z), standard deviation of unbiased gene diversity (UH_z s.d.), observed heterozygosity (OH_z), standard deviation of observed heterozygosity (OH_z s.d.), number of allele *per locus* (n_A), standard deviation of number of allele *per locus* (n_A s.d.), number of effective alleles (n_e), standard deviation of the number of effective alleles (n_e s.d.), and polymorphic index content values (PIC), are shown. The maximum values are written in bold.

Location	N	UH_z	UH_z s.d.	OH_z	OH_z s.d.	n_A	n_A s.d.	n_e	n_e s.d.	PIC
1	1	0.40	0.24	0.40	0.22	1.40	0.55	1.40	0.24	0.15
2	3	0.36	0.15	0.13	0.09	1.80	0.84	1.63	0.29	0.25
3	2	0.43	0.18	0.60	0.15	1.80	0.84	1.73	0.32	0.26
4	3	0.49	0.13	0.47	0.13	2.20	0.84	1.87	0.25	0.34
5	8	0.45	0.15	0.60	0.08	**3.20**	2.77	**2.58**	0.98	0.37
6	4	0.40	0.17	0.60	0.11	2.20	1.64	2.00	0.54	0.29
7	3	0.23	0.14	0.33	0.12	1.40	0.55	1.36	0.22	0.14
8	9	0.45	0.15	0.56	0.07	3.00	2.35	2.39	0.72	0.38
9	9	0.41	0.17	0.51	0.07	3.00	2.35	2.32	0.72	0.34
10	10	0.43	0.15	0.58	0.07	2.80	1.92	2.34	0.74	0.36
11	10	0.40	0.17	0.54	0.07	2.60	2.07	2.34	0.76	0.34
12	8	0.43	0.16	**0.63**	0.08	2.80	1.92	2.20	0.59	0.35
13	6	0.40	0.17	0.47	0.09	2.40	1.52	2.01	0.47	0.32
14	2	0.50	0.21	0.60	0.15	2.20	1.30	2.13	0.56	0.33
15	2	0.53	0.14	0.50	0.16	2.00	0.71	1.85	0.27	0.32
16	3	0.57	0.12	0.53	0.13	2.80	1.30	2.32	0.58	**0.42**
17	1	**0.60**	0.24	0.60	0.22	1.60	0.55	1.60	0.24	0.23
All samples	84	0.45	0.15	0.53	0.02	3.80	2.68	2.75	2.14	0.41

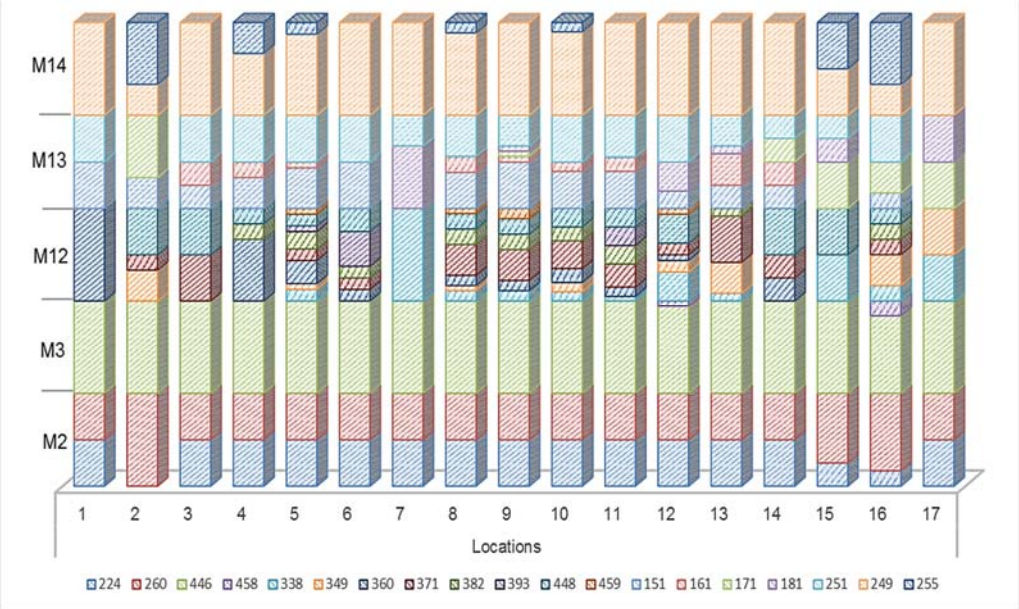

Figure 3.16 Allelic frequencies pattern *per* location based on SSR *loci*. Alleles of microsatellites *loci* (M2, M3, M12, M13 and M14) are shown for each location. All alleles are colored and distributed by their frequency within the scale of each *locus*. Each location shows a unique allelic frequency pattern for each location.

3.1.5.3 Population and genetic structure

3.1.5.3.1 Sample level

The population structure based on non-spatial Bayesian clustering method as implemented in STRUCTURE differentiated the samples into two groups. Samples from the same locations shared similar membership coefficients in inferred clusters. From K = 2 to K = 10, K= 2 was chosen by applying the criteria of highest average of the “estimated ln Prob (K)” and the lowest “variance of ln likelihood” (Table 3.24; Fig. 3.17). In the bar plot produced by STRUCTURE samples were separated into group A (red) and group B (green) (Fig. 3.18).

Table 3.23 Allele frequency *per* locus in the 17 studied locations based on five microsatellite markers. Alleles length in bp (A), total frequency of each allele (T), allele count *per* locus (IC) and allele count *per* location (AC), are shown. Highest values are written in bold.

SSR *Loci*	A	Location																	T	IC
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17		
M2	224	50	0	50	50	50	50	50	50	50	50	50	50	50	50	25	16.7	50	46.4	78
	260	50	100	50	50	50	50	50	50	50	50	50	50	50	50	75	83.3	50	53.6	**90**
M3	446	100	100	100	100	100	100	100	100	100	100	100	93.8	100	100	100	83.3	100	98.8	**166**
	458	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	6.3	0.0	0.0	0.0	16.7	0.0	1.2	2
	338	0.0	0.0	0.0	0.0	12.5	0.0	100	11.1	11.1	10	5	31.3	8.3	0.0	50	16.7	50	14.9	25
	349	0.0	33.3	0.0	0.0	6.3	0.0	0.0	5.6	0.0	10	0.0	12.5	33.3	0.0	0.0	33.3	50	8.9	15
	360	100	0.0	0.0	66.7	25	12.5	0.0	11.1	11.1	15	10	6.3	0.0	25	0.0	0.0	0.0	13.1	22
M12	371	0.0	16.7	50	0.0	12.5	12.5	0.0	33.3	33.3	30	25	12.5	50	25	0.0	16.7	0.0	23.2	39
	382	0.0	0.0	0.0	16.7	18.8	12.5	0.0	16.7	16.7	15	20	0.0	8.3	0.0	0.0	16.7	0.0	11.9	20
	393	0.0	0.0	0.0	0.0	6.3	37.5	0.0	0.0	0.0	0.0	20	0.0	0.0	0.0	0.0	0.0	0.0	4.8	8
	448	0.0	50	50	16.7	12.5	25	0.0	16.7	16.7	20	20	31.3	0.0	50	50	16.7	0.0	20.2	34
	459	0.0	0.0	0.0	0.0	6.3	0.0	0.0	5.6	11.1	0.0	0.0	6.3	0.0	0.0	0.0	0.0	0.0	3	5
	151	50	33.3	25	33.3	43.8	50	0.0	38.9	50	40	40	18.8	25.0	25	0.0	16.7	0.0	33.9	**57**
	161	0.0	0.0	25	16.7	6.3	0.0	0.0	16.7	5.6	10	15	0.0	33.3	25	0.0	0.0	0.0	10.1	17
M13	171	0.0	66.7	0.0	0.0	0.0	0.0	0.0	0.0	5.6	0.0	0.0	0.0	0.0	25	50	33.3	50	6.5	11
	181	0.0	0.0	0.0	0.0	0.0	0.0	66.7	0.0	5.6	0.0	0.0	31.3	8.3	0.0	25	0.0	50	7.7	13
	251	50	0.0	50	50	50	50	33.3	44.4	33.3	50	45	50	33.3	25	25	50	0.0	41.7	**70**
M14	249	100	33.3	100	66.7	87.5	100	100	88.9	100	90	100	100	100	100	50	33.3	100	89.3	**150**
	255	0.0	66.7	0.0	33.3	12.5	0.0	0.0	11.1	0.0	10	0.0	0.0	0.0	0.0	50	66.7	0.0	10.7	18
AC	19	7	8	9	10	**15**	11	7	14	**15**	13	13	14	12	11	9	13	8	-	164

Table 3.24 Bayesian structure method output for microsatellite *loci* of 84 samples in 17 locations. Estimated ln Prob (LnP), mean values of ln likelihood (MLn) and variance of ln likelihood (VLn) for K = [2 – 9], are shown. Each K-value was run three times under the same parameters.

K	LnP	MLn	VLn	K	LnP	MLn	VLn
2	-717.2	-675.3	83.9	6	-775.2	-659.6	231.2
2	-727.8	-676.1	103.2	6	-911.8	-639.3	544.9
2	-718.4	-676.4	84.2	6	-748.9	-737.8	22.2
Average	**-721.1**	**-675.9**	**90.4**	**Average**	**-811.97**	**-678.9**	**266.1**
3	-769.8	-667.2	205.2	7	-781.1	-656.8	248.6
3	-949.2	-706.4	485.6	7	-819.5	-638.3	362.4
3	-736.8	-662.2	149.3	7	-1092.4	-651.3	882.3
Average	**-818.6**	**-678.6**	**280.0**	**Average**	**-897.7**	**-648.8**	**497.8**
4	-830.8	-656.5	348.8	8	-970.7	-661.9	-617.5
4	-863.3	-674	378.5	8	-762	-645.3	233.4
4	-893.8	-655.6	476.4	8	-2021.7	-690.9	2661.5
Average	**-862.6**	**-662.03**	**401.2**	**Average**	**-1251.5**	**-666.03**	**759.1**
5	-732.3	-644.1	136.5	9	-744.2	-738.6	11.3
5	-795.7	-634.5	332.4	9	-722.2	-638.3	167.7
5	-743.1	-738.9	8.4	9	-737.2	-636.9	200.7
Average	**-757.0**	**-672.5**	**159.1**	**Average**	**-734.5**	**-671.3**	**126.6**

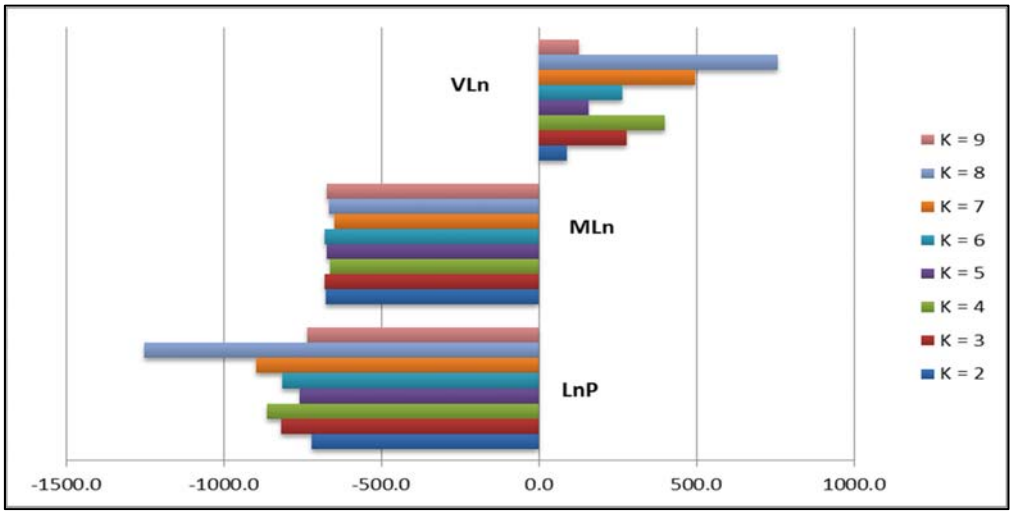

Figure 3.17 Graphical plotting of STRUCTURE output data based on SSR *loci*. Estimated ln probability (LnP), mean values of ln likelihood (MLn) and variance of ln likelihood (VLn) for K = [2 – 9], are shown. K = 2 shows the lowest VLn, moderate MLn and highest LnP.

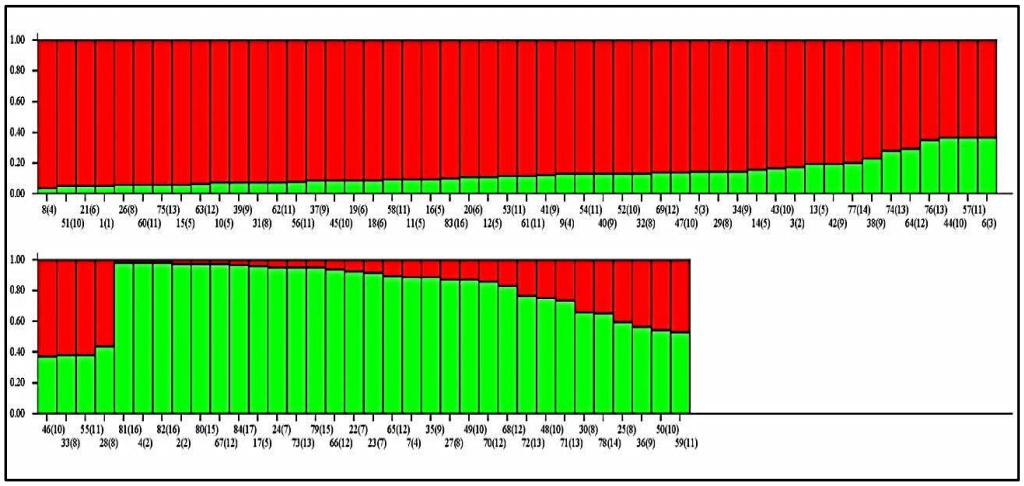

Figure 3.18 STRUCTURE output bar plot graph of K = 2 for 84 samples in 17 locations based on SSR *loci*. Samples are ordered by group assignment, group A (red) and group B (green). Locations are written between brackets.

3.1.5.3.2 Location level

When locations 1 and 17 were excluded, a high gene flow (N_m) of 2.00 based on wright's fixation index (F_{ST}) was calculated by POPGENE software.

An AMOVA test was estimated by ARLEQUIN based on the pairwise differences, to test the genetic structure by comparing results from various groupings and choosing the structure that maximizes the among locations variance. F_{ST} was equal to 0.20, partitioned into a major genetic variation originated within locations, accounting for 79 % of the total variation, while only 21 % of the genetic variation occurred among locations (Table 3.25).

Table 3.25 Genetic differentiation estimated by AMOVA of *Funaria hygrometrica* based on microsatellite data. The source of variance (among and within locations), the degree of freedom (df), the sum of squares (SS), the variance components and the percentage of variation, are shown.

Source of variance	df	SS	Variance components	Percentage of variation
Among locations	14	49.91	0.38	21 %
Within locations	67	99.79	1.49	79 %
Total	81	149.70	1.87	

F_{ST} = 0.20 (p-value = 0.00)

3.1.6 AFLP

3.1.6.1 Amplification, scoring and readability

PCR amplification was successful for 12 pairs of primers. The rate of amplification was up to 195 peaks *per* primer pair (e.g. sample 6, ACA – CAA; Fig. 3.19 A, B). Peak analysis and automated band scoring were successful and quality tests showed adequate quality. Band scoring for each primer pair were between 150 bp – 600 bp (Table 3.26).

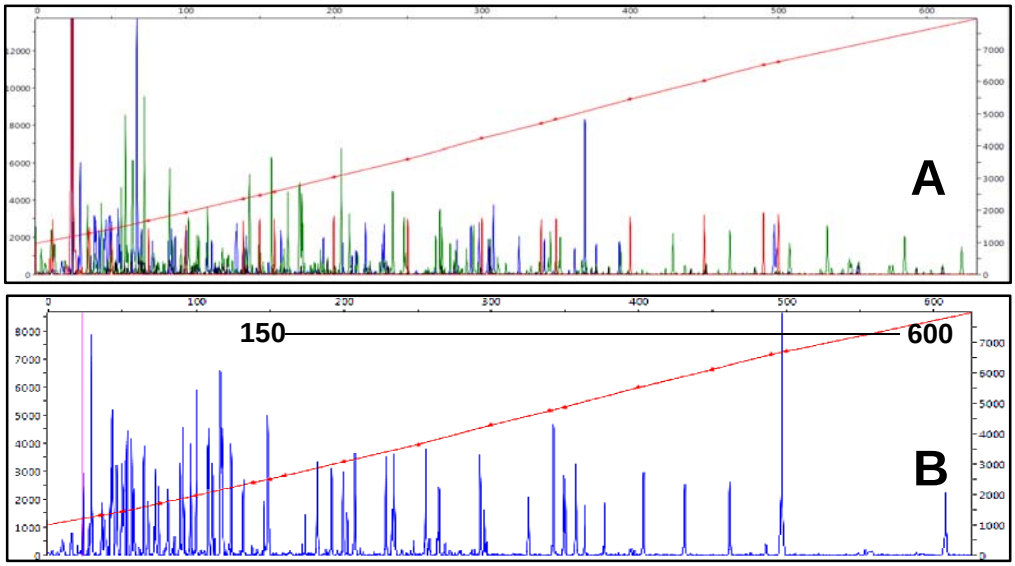

Figure 3.19 PEAKSCANNER chromatogram of AFLP primer pairs for sample 6. A: peaks of ACA-CAA (blue), AGG-CAA (green) and ATA-CAA (black). B: peaks of primer pair ACA-CAA (blue), only peaks of 150 – 600 bp were used for band scoring.

3.1.6.2 Population data, total and within locations genetic variation

Plotting randomly selected 1 584 *loci* against their calculated genetic diversity reached a plateau approximately at mean genetic diversity of 1.00. This confirmed that the current scored bands were sufficient and no more *loci* were needed for better estimation of genetic diversity (Fig. 3.20).

Table 3.26 Primer pairs amplification and scoring results. Primer code (PC), number of peaks (NP), scored bands (SB), quality check (QC), total number of bands (T) and mean of quality check (M), are shown.

PC	NP	SB	QC	PC	NP	SB	QC
ACA-CAA	40 - 167	304	0.70	ATA-CAT	1 – 36	79	0.60
AGG-CAA	60 - 195	398	0.80	ACA-CTA	8 – 93	211	0.70
ATA-CAA	9 - 64	185	0.60	AGG-CTA	33 - 105	246	0.90
ACA-CAC	21 - 118	287	0.80	ATA-CTA	2 – 48	157	0.80
AGG-CAC	27 - 105	240	0.90	ACA-CTC	5 – 47	136	0.80
ATA-CAC	21 - 65	189	0.60	AGG-CTC	13 - 67	193	0.80
ACA-CAT	1 - 53	143	0.60	ATA-CTC	1 – 46	95	0.60
AGG-CAT	3 - 74	176	0.70			T = 3 057	M = 0.72

A total of 3 057 bands were scored from all primer pairs for all the 84 samples. 3 056 bands were polymorphic, only one was monomorphic; 1 411 were rare bands with frequency less than 5 % and 61 bands had frequencies higher than 95 %. Only 1 584 *loci* were retained at 5 % (i.e. after removing all bands with frequency less than 5 %). The mean fragment size was ~ 321 bp with a standard deviation of ~ 121 bp. A weak negative correlation was found between fragment sizes and frequencies (r = -0.19; p-value 0.00). The mean number of band presence *per* sample was ~ 607 bands with a standard deviation of 93 bands (Table 3.27). Sample 3 of location 2 showed the lowest band frequency *per* sample (0.08). In contrast, sample 82 of location 16 showed the highest band frequency *per* sample (0.26) (Fig. 3.21).

The effective number of alleles (n_e) for all samples combined together was 1.39. The expected heterozygosity under Hardy-Weinberg assumption (H_e) for all samples combined together was 0.21. Locations 1 and 17 were excluded when the estimated parameters between locations were compared for having one sample. The maximum and minimum number of polymorphic bands (PB) were 1 715 (1 250 at 5 %) bands for location 9 and 237 (211 at 5 %) bands for location 3, respectively. The maximum and minimum percentages of polymorphic bands at 5 % (PLP at 5 %) were 78.9 % for location 9 and 13.3 % for location 3, respectively. The maximum and minimum private bands *per* location (PrB) were 153 bands for location 9 and

four bands for location 3, respectively. No fixed bands were detected except for location 7, where seven private bands were fixed. Location 3 yielded the lowest n_e (1.13) and the lowest H_e (0.13) when F_{IS} = 1. However, location 14 scored the highest n_e (1.49), but location 15 has the highest H_e (0.44) (Table 3.27).

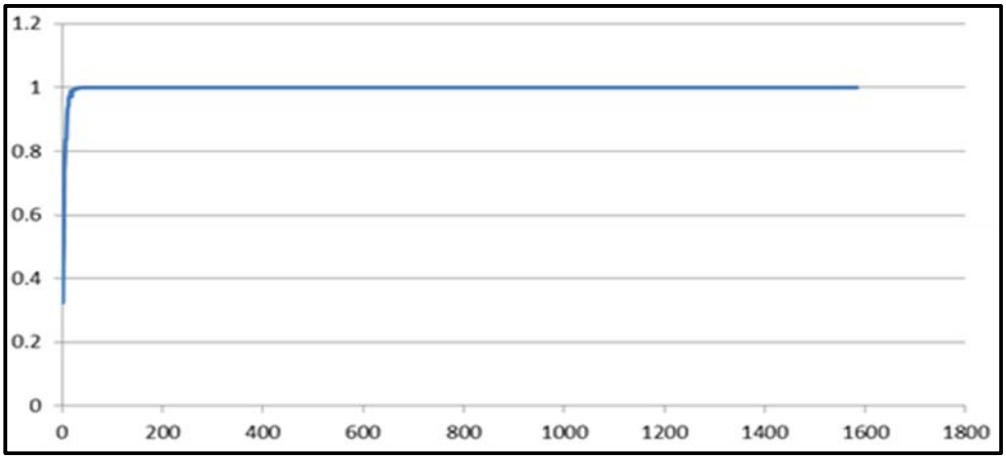

Figure 3.20 Number of randomly selected 1 584 AFLP *loci* plotted against mean genetic diversity, where a plateau had been reached approximately near 1.00.

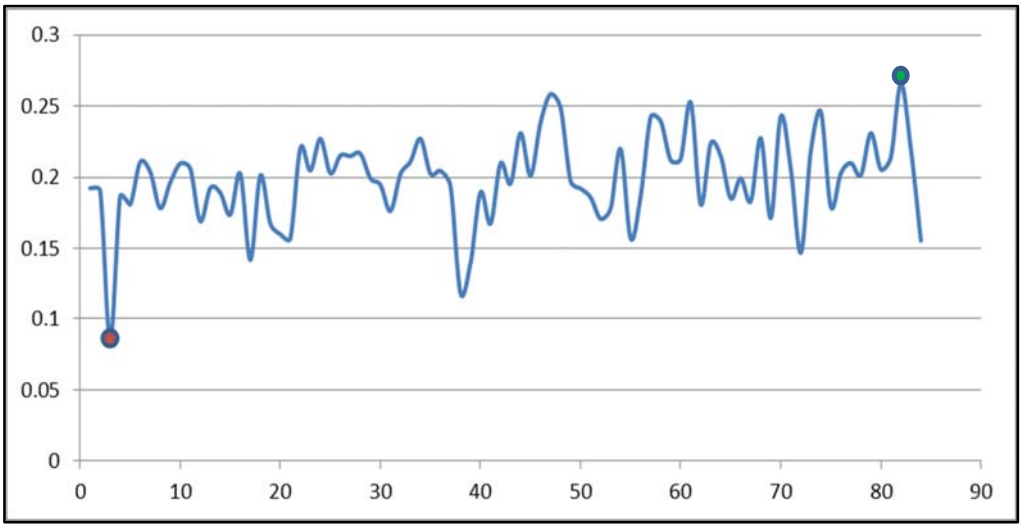

Figure 3.21 Frequency of AFLP *loci per* sample. Frequency was measured for 84 samples in 17 locations and ranged from 0.08 (red dot) to 0.26 (green dot).

Table 3.27 Genetic variation of AFLP molecular markers *per* location. Sample locations (Loc), sample size (N), number of polymorphic bands (PB), number of polymorphic bands at 5 % (PB at 5 %), proportion of polymorphic *loci* at the 5 % level - expressed as a percentage (PLP at 5 %), number of private bands (PrB), number of fixed private bands (FB), effective number of alleles (n_e), effective number of alleles standard deviation (n_e s.d.), expected heterozygosity under Hardy-Weinberg (H_e), and expected heterozygosity standard deviation (H_e s.d.), are shown. The highest and the lowest values are written in bold.

Loc	N	PB	PB at 5 %	PLP at 5 %	PrB	FB	n_e	n_e s.d.	H_e	H_e s.d.
1	1	0.00	0.00	0.00	10	10	1.00	0.00	0.00	0.00
2	3	663	540	34.1	27	0	1.34	0.47	0.23	0.32
3	2	**237**	**211**	**13.3**	**4**	0	**1.13**	**0.34**	**0.13**	**0.34**
4	3	612	449	28.3	44	0	1.28	0.45	0.19	0.30
5	8	751	635	40.1	25	0	1.40	0.49	0.16	0.21
6	4	636	578	36.5	15	0	1.37	0.48	0.19	0.26
7	3	522	416	26.3	19	**7**	1.26	0.44	0.18	0.29
8	9	1097	785	49.6	92	0	1.50	0.50	0.17	0.20
9	9	**1715**	**1250**	**78.9**	**153**	0	1.79	0.41	0.29	0.19
10	10	949	759	47.9	38	0	1.48	0.50	0.18	0.21
11	10	1074	842	53.2	47	0	1.53	0.50	0.19	0.21
12	8	1245	1097	69.3	32	0	1.69	0.46	0.31	0.23
13	6	1276	1089	68.8	37	0	1.69	0.46	0.29	0.21
14	2	594	467	29.5	34	0	1.30	0.46	0.29	0.46
15	2	798	698	44.1	15	0	1.44	0.50	**0.44**	**0.50**
16	3	953	775	48.9	41	0	**1.49**	**0.50**	0.33	0.33
17	1	0.00	0.00	0.00	16	16	1.00	0.00	0.00	0.00
All samples	-	607	-	-	38.17	-	1.39	0.22	0.21	0.11
Total	84	3056	1584	-	649	-	-	-	-	-

3.1.6.3 Population and genetic structure

3.1.6.3.1 Haplotype and sample level

An approximated minimum spanning tree (Fig. 3.22) of all individuals showed a separation of the individuals into three groups, while intermediate individuals appeared in between. Group A contained the highest number of samples (60) and group B had 12 samples. Both were separated by intermediate samples. The in between samples were 36, 38, 39, 41 and 77 all from location 9 except for sample 77 from location 12. Group C included 6 samples (2, 3, 4, 80, 81 and 82), and was connected to group A through sample 78.

An analysis of the population structure based on non-spatial Bayesian clustering as implemented in STRUCTURE was carried out. From K = 2 to K = 10, K = 4 and K= 5 were chosen by applying the criteria of highest average estimated ln probability score and lowest variance. Both were re-estimated again for five more times. Based on the previously mentioned criteria, K= 4 was determined as the best K value that reflects the population structure (Table 3.28; Fig. 3.23). Samples from the same location shared similar membership coefficients in inferred clusters. In a comparison with the minimum spanning tree, group A (represented in green and blue) showed some internal structure in form of three sub-groups, but group B was as same as it was shown in the minimum spanning tree (represented in yellow) (Fig. 3.24).

3.1.6.3.2 Location level

When locations 1 and 17 were excluded, gene flow (N_m) of 1.88 based on Wright's fixation index (F_{ST}) and calculated by AFLP-SURV software.

An AMOVA test was estimated by ARLEQUIN based on the pairwise differences, to test the genetic structure by comparing the results from various groupings and choosing the structure that maximizes the among locations variance. The F_{ST} was equal to 0.21, partitioned into a major genetic variation originated within locations, accounting for 79 % of the total variations, while 21 % of the genetic variation occurred among locations (Table 3.29).

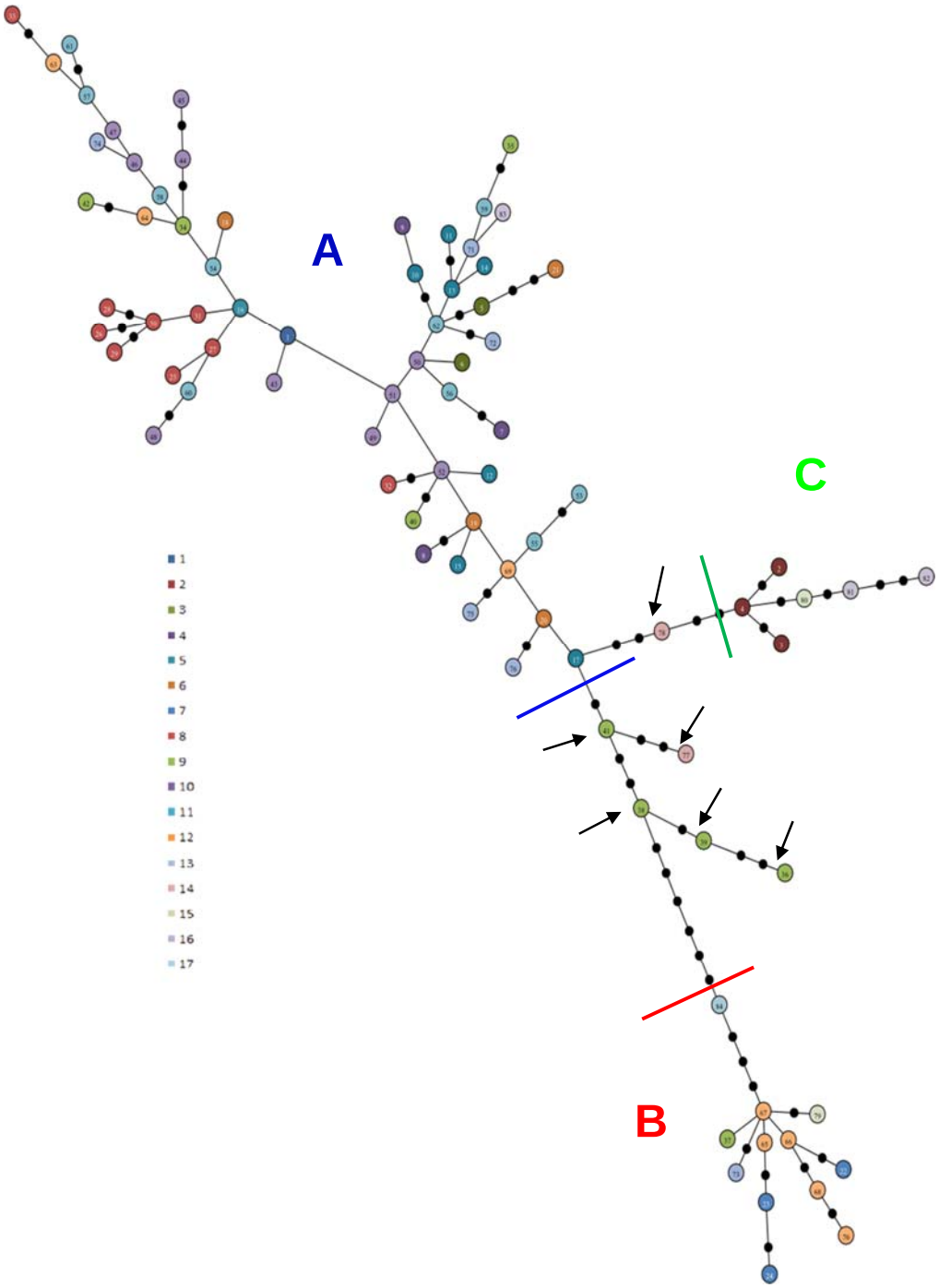

Figure 3.22 Minimum spanning tree of AFLP-generated 84 haplotypes from 84 samples in 17 locations. Group A is determined by a blue line while group C is separated by a green line. Group B is separated by a red line. All haplotypes possess the same size and the same sample frequency (equal to 1), and are colored according to their sampling location. Black arrows point to intermediate samples in between the shown groups.

Table 3.28 Bayesian population structure method applied on 1 584 AFLP *loci* of 84 samples in 17 locations (STRUCTURE). The ln probability (LnP), mean values of ln likelihood (MLn) and variance of ln likelihood (VLn) for K = {4, 5}, are shown. Each K-value was run five times under the same parameters.

K	LnP	MLn	VLn
4	-41344	-39929	2831
4	-385078	-40371	689415
4	-46829	-39953	13752
4	-41361	-39937	2848
4	-41352	-39941	2823
4	-41509	-39932	3154
Average	**-99579**	**-40010**	**119137**
5	-48784	-38600	20367
5	-397529	-39072	716915
5	-69555	-38684	61743
5	-412681	-38576	5384
5	-68595	-38663	59863
5	-41084	-38579	5010
Average	**-173038**	**-38696**	**144880**

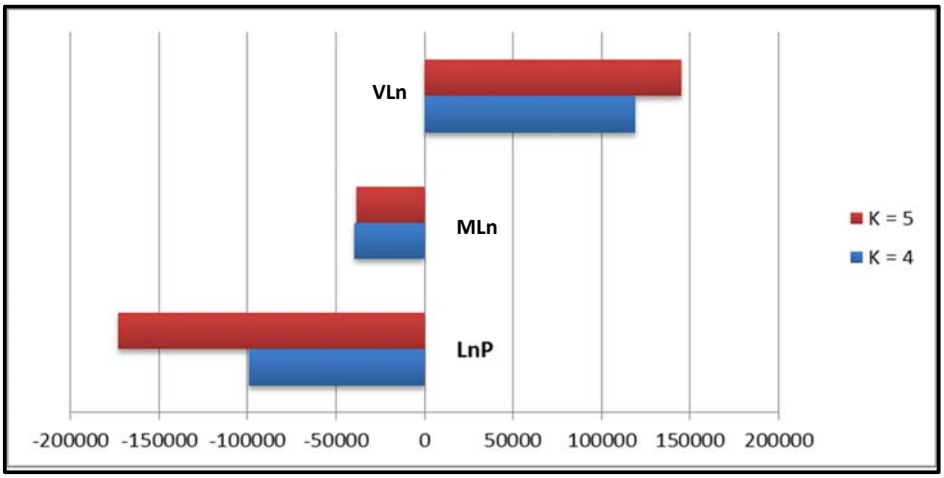

Figure 3.23 Graphical plotting of STRUCTURE output data based on AFLP *loci*. Estimated ln probability (LnP), mean values of ln likelihood (MLn) and variance of ln likelihood (VLn) for K = {4, 5}. K = 4 shows the highest LnP and lowest VLn.

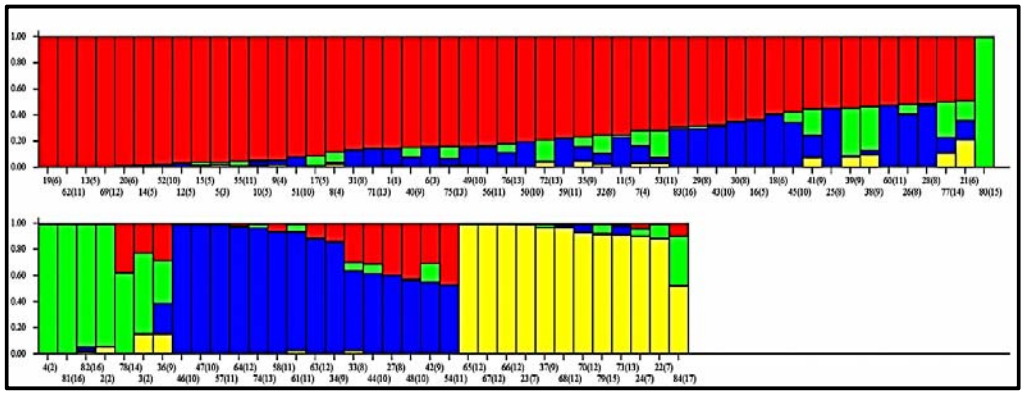

Figure 3.24 AFLP marker-based structure output bar plot graph of K = 4, for 84 samples in 17 locations. Samples are ordered by group assignment. Red and blue correspond to group A, green to group C and yellow to group B from the minimum spanning tree. Locations are indicated between brackets.

Table 3.29 Genetic differentiation through AMOVA of *Funaria hygrometrica* based on AFLP data. The source of variance (among and within locations), the degree of freedom (df), the sum of squares (SS), the variance components and the percentage of variation, are shown.

Source of variance	df	SS	Variance components	Percentage of variation
Among locations	14	6113.23	48.40	21 %
Within locations	67	11927.78	178.02	79 %
Total	81	18041.01	226.43	

F_{ST} = 0.21 (p-value = 0.00)

3.1.7 MULTIPLE COMPARISON AND DATA COMBINING

A comparison scheme between the different sequenced DNA regions (ITS1, ITS2, cpDNA and mtDNA) and the fingerprinting techniques (SSR and AFLP techniques) is given in Table 3.30. Each region and technique show different levels of variations and evolution, but common aspects can be found at different comparative levels.

Generally, haplotypes from the six molecular markers can be grouped into two groups. Group A was colored in blue and group B in red. To track the haplotype distributions in all sampled locations, the colored table demonstrates the haplotype state of each sample and its correspondence to each group. Mainly the samples of low and middle altitudes (up to 1 300 m) except for location 2, showed a homogenate state (i.e. they belong to group A or group B for all studied markers). Within location 9, only sample 37 showed a homogenate state of a different group than the other samples. On higher altitudes the samples started to show non-homogenate and/or homogenate states from both groups within the same location.

Through the table, samples 2 to 4, 60, 63, 65, 67, 70, 73, 76, 78 to 82 and 84 were identified as non-homogenate samples. Samples 79, 82 and 84 were non-homogenate only on maternal lines level between chloroplast and mitochondrial haplotypes. Samples 60, 63, 70, 76 and 80 were non-homogenate only on nuclear level between ITS1 and ITS2. While samples 73 and 81 were non-homogenate on both levels. Samples 2 to 4, 65, 67, 70 and 78 were homogenate on maternal lines level and nuclear level, but each level corresponds to a different group.

AFLP marker reflected similar results compared with the sequenced *loci* in respect to sample assignments to groups A or B. However, in the case of the microsatellite marker there are some remarkable differences, as samples 7, 17, 27 and 49 the assignment based on microsatellites contradicts all other markers.

Table 3.30 Comparative table between all markers applied in the current study. For each sample and marker, the haplotype number is indicated. Blue cells represent haplotypes belonging to group A and red cells haplotypes belonging to group B. The table shows the homogeneity state of each sample.

Location	Sample	ITS1	ITS2	cpDNA	mtDNA	SSR	AFLP	Location	Sample	ITS1	ITS2	cpDNA	mtDNA	SSR	AFLP
1.1	1	1	1	1	1			10.1	43	3	1	16	1		
2.1	2	2	2	2	2			10.2	44	6	1	10	1		
2.2	3	2	2	2	2			10.3	45	3	1	1	1		
2.3	4	2	2	2	2			10.4	46	5	1	17	1		
3.1	5	3	1	1	1			10.5	47	15	1	1	1		
3.2	6	4	1	3	1			10.6	48	3	1	1	1		
4.1	7	3	1	1	1			10.7	49	1	1	18	1		
4.2	8	3	1	4	1			10.8	50	3	1	19	1		
4.3	9	3	1	1	3			10.9	51	17	1	1	1		
5.1	10	5	1	1	1			10.10	52	3	1	1	7		
5.2	11	6	1	1	1			11.1	53	3	1	1	1		
5.3	12	7	1	1	1			11.2	54	3	1	1	1		
5.4	13	8	1	1	1			11.3	55	15	1	1	1		
5.5	14	8	1	-	1			11.4	56	3	1	1	1		
5.6	15	3	1	-	1			11.5	57	3	1	1	1		
5.7	16	8	1	5	1			11.6	58	3	1	1	1		
5.8	17	3	1	-	1			11.7	59	3	1	1	1		
6.1	18	3	1	1	1			11.8	60	9	1	11	1		
6.2	19	3	1	1	1			11.9	61	18	1	1	1		
6.3	20	3	1	6	1			11.10	62	15	1	1	1		
6.4	21	3	1	1	1			12.1	63	20	5	1	1		
7.1	22	9	3	7	2			12.2	64	5	1	1	1		
7.2	23	9	4	7	2			12.3	65	3	1	20	2		
7.3	24	9	3	7	2			12.4	66	9	3	20	5		
8.1	25	10	1	-	1			12.5	67	3	1	20	6		
8.2	26	3	1	8	1			12.6	68	9	3	20	2		
8.3	27	11	1	9	1			12.7	69	3	6	21	1		
8.4	28	12	1	1	1			12.8	70	19	7	20	2		
8.5	29	3	1	10	1			13.1	71	3	1	1	1		
8.6	30	3	1	11	1			13.2	72	3	1	1	4		
8.7	31	6	1	1	1			13.3	73	9	1	22	1		
8.8	32	13	1	12	1			13.4	74	3	1	23	1		
9.1	33	3	1	13	1			13.5	75	21	1	1	8		
9.2	34	3	1	14	1			13.6	76	9	1	1	1		
9.3	35	14	1	1	1			14.1	77	3	2	1	1		
9.4	36	3	1	1	1			14.2	78	6	2	24	9		
9.5	37	9	3	15	2			15.1	79	3	2	1	10		
9.6	38	15	1	1	1			15.2	80	22	2	25	11		
9.7	39	3	1	13	1			16.1	81	9	1	2	2		
9.8	40	16	1	1	1			16.2	82	9	3	2	2		
9.9	41	3	1	1	1			16.3	83	15	1	2	1		
9.10	42	14	1	1	1			17.1	84	9	3	1	2		

The test for linkage disequilibrium of the four different *loci* used for sequencing found significant values of disequilibrium for all combinations of *loci* with the exception of the pair ITS1 – cpDNA. *Loci* in disequilibrium are not independent of another but rather associated in a non-random manner. In other words, certain combinations of haplotypes of different *loci* are more frequent than one would expect in a random process while others are less frequent. Only the combination of the ITS1 spacer with the cpDNA region is slightly above the significance level of 0.05 with a value of 0.06 (Table 3.31).

When the populations 1 to 11 (below 1 300 m a.s.l.) were studied for possible linkage disequilibrium, all combinations of *loci* were in disequilibrium with p-values of 0.00. But when the populations 12 to 17 were considered, only ITS1 and ITS2 show a non-random association, which can be expected because the two *loci* are only separated by the short 5.8S rRNA gene and arranged in tandem repeat units (Table 3.31).

Table 3.31 Table of linkage disequilibrium of the four genotyped *loci* for all samples combined together, for locations below 1 300 m a.s.l. (locations 1 to 11) and for locations above 1 300 m a.s.l. (locations 12 to 17). P-values < 0.05 are written in bold.

Loci	**All samples**				**Below 1 300 m a.s.l.**				**Above 1 300 m a.s.l.**			
	ITS1	ITS2	cpDNA	mtDNA	ITS1	ITS2	cpDNA	mtDNA	ITS1	ITS2	cpDNA	mtDNA
ITS1												
ITS2	**0.00**				**0.00**				**0.04**			
cpDNA	0.06	**0.00**			**0.00**	**0.00**			0.49	0.50		
mtDNA	**0.00**	**0.00**	**0.00**		**0.00**	**0.00**	**0.00**		0.31	0.21	0.22	

PART 3.2

Adaptation to

Environmental Factors

3.2.1 LOCI UNDER SELECTION

3.2.1.1 Detection of positive selection signatures

Markers under selection in a variable environment are supposed to show higher fixation indices than neutral markers as a consequence of the stronger differentiation between the populations for these markers. The AFLP data set of *Funaria hygrometrica* was analyzed for outlier *loci* detection by using the MCHEZA program on the 17 locations divided into four altitudinal groups (see point 2.1.2, pp. 56). Across the 16 pairwise analysis between the four groups, 12 out of 1 584 *loci* (0.7 %) were identified as outlier *loci* under directional selection at the 99.5 % confidence level (Fig. 3.25). The 12 outliers *loci* were the followings: numbers 375, 414, 871, 1 021, 1 135, 1 143, 1 254, 1 476, 1 907, 2 524 and 2 589. The mentioned *loci*, constantly appeared as outlier *loci* among the four groups in each run. Un-constant *loci* were considered as false positive (i.e. *loci* detected because of the 5% type I error). None of the 12 *loci* showed a complete bias or has been attached to a particular population, while they all were polymorphic in compared groups. The mean F_{ST} value for the 12 *loci* was 0.67.

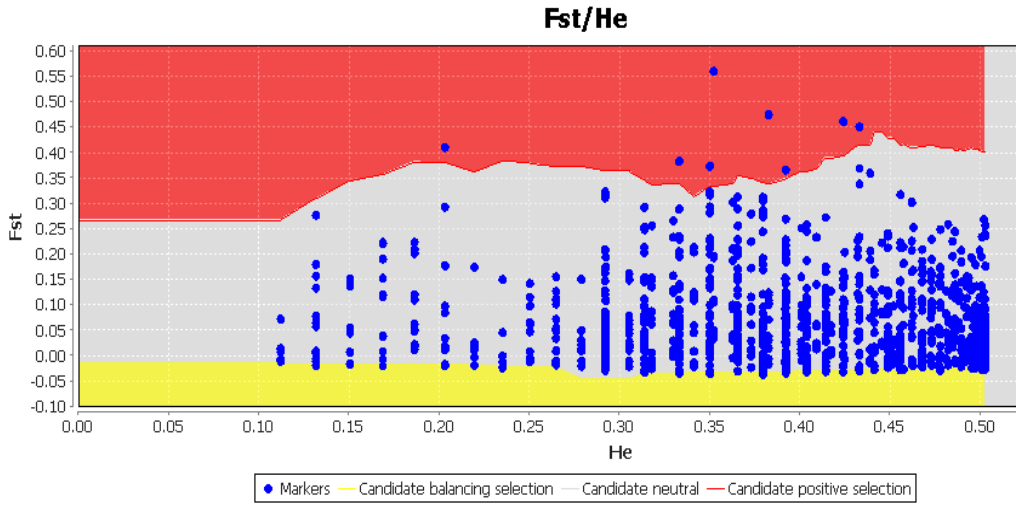

Figure 3.25 Graphical plot of F_{ST} values against heterozygosity (H_e) for each of the 1 584 AFLP *loci*. Each dot indicates an AFLP *locus*. The lower (yellow), and higher (red) zones represent the 0.5 % and 99.5 % confidence intervals, respectively. *Loci* in the red zone above the 99.5 % are regarded as outlier *loci*.

3.2.1.2 Analysis of molecular variance (AMOVA)

An AMOVA test was used to measure the changes in the pairwise differentiation of the F_{ST} for the AFLP neutral *loci* dataset in comparison with the F_{ST} of AFLP 12 outlier *loci* dataset. F_{ST} increased from 0.19 in AFLP neutral *loci* to 0.62 in AFLP 12 outlier *loci*. The percentage of variation due to differences within locations decreased from 81 % in AFLP neutral *loci* to 38 % in AFLP 12 outlier *loci*. In the AFLP 12 outliers *loci* dataset, F_{ST} partitions were 36 % between groups, 26 % among locations between groups and 38 % within locations (Tables 3.32 and 3.33; Fig. 3.26).

Table 3.32 Genetic differentiation through AMOVA of *Funaria hygrometrica* based on the AFLP neutral *loci* dataset, where no groups were assumed. The source of variance (among and within locations), the degree of freedom (df), the sum of squares (SS), the variance components and the percentage of variation, are shown.

Source of variance	df	SS	Variance components	Percentage of variation
Among locations	14	6657.62	49.80	19 %
Within locations	67	14032.85	209.44	81 %
Total	81	18041.01	226.43	

F_{ST} = 0.19 (p-value = 0.00)

Table 3.33 Genetic differentiation through AMOVA of *Funaria hygrometrica* based on AFLP 12 outliers *loci* dataset, after four groups were assumed. The source of variance (among and within locations), the degree of freedom (df), the sum of squares (SS), the variance components and the percentage of variation, are shown.

Source of variance	df	SS	Variance components	Percentage of variation
Between groups	3	40.43	0.56	36 %
Among locations within groups	11	29.91	0.40	26 %
Within locations	67	39.86	0.59	38 %
Total	81	110.22	1.56	

F_{ST} = 0.62 (p-value = 0.00)

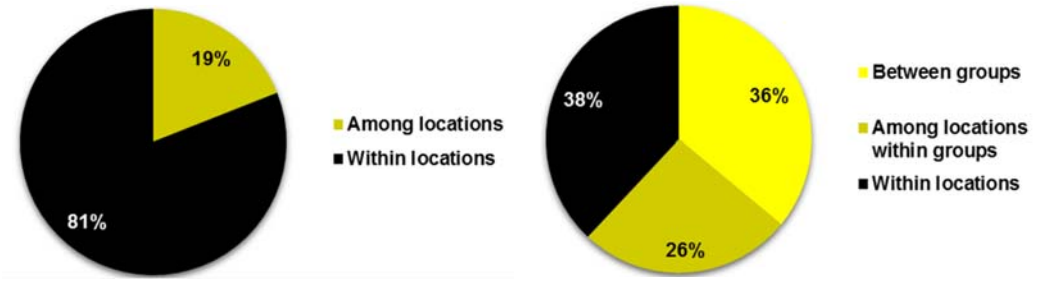

Figure 3.26 F_{ST} partitions based on AMOVA test of the neutral AFLP *loci* (left) and the AFLP 12 outliers *loci* (right).

3.2.2 CORRELATION TESTS

A PCA of bioclimatic variables (see point 2.2.6, pp. 84) showed clustering in three sets and five single variables (Fig. 3.27; Table 3.34). Correlation tests between each set and each single variables showed that singles 1 and 5 were highly correlated to set 2, single 2 was highly correlated to set 3 and single 3 was highly correlated to set 1 (Table 3.35). According to these data, finally three sets and 1 single variable (single 4) were used to generate a separate Euclidean distance matrix.

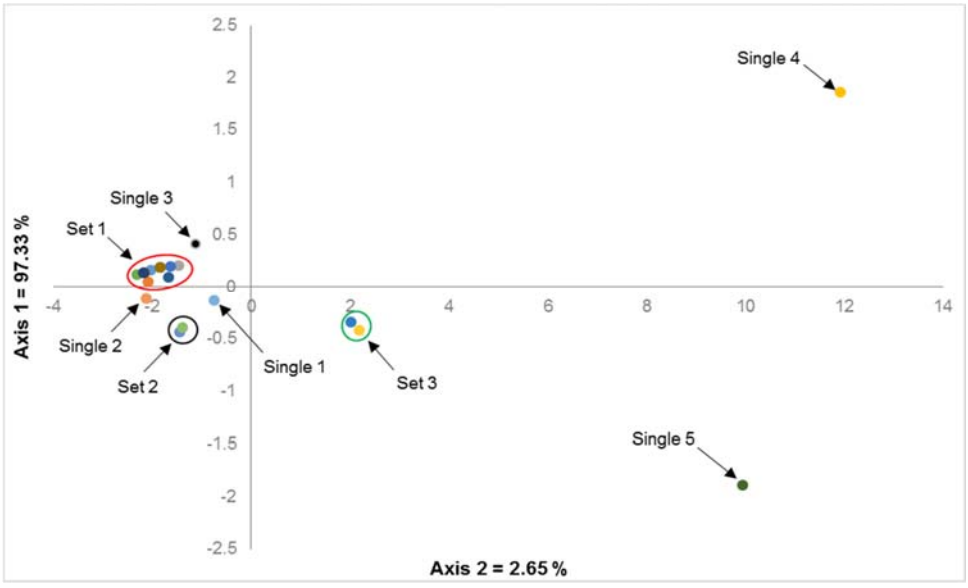

Figure 3.27 PCA plot for 19 bioclimatic variables explained by two axes. Axis 1 = 97.33 % and axis 2 = 2.65 %. Three sets and five singles were defined.

Table 3.34 Bioclimatic variables clustering according to PCA analysis.

Bioclimatic variable	Set	Bioclimatic variable	Set
Annual mean temperature	Set 1	Precipitation of wettest quarter	Set 2
Mean monthly temperature range	Set 1	Precipitation of coldest quarter	Set 2
Isothermality	Set 1	Precipitation of driest quarter	Set 3
Maximal temperature of warmest month	Set 1	Precipitation of warmest quarter	Set 3
Minimal temperature of coldest month	Set 1	Precipitation of wettest month	Single 1
Temperature annual range	Set 1	Precipitation of driest month	Single 2
Mean temperature of wettest quarter	Set 1	Precipitation seasonality	Single 3
Mean temperature of driest quarter	Set 1	Temperature seasonality	Single 4
Mean temperature of warmest quarter	Set 1	Annual precipitation	Single 5
Mean temperature of coldest quarter	Set 1		

Table 3.35 Correlation test between each set and each single of bioclimatic variables clustered by PCA analysis. Highest values for each single variable are written in bold.

Variable	Set 1	Set 2	Set 3	Single 1	Single 2	Single 3	Single 4	Single 5
Set 1	1	0.95	0.86	0.95	0.86	**0.96**	0.71	0.95
Set 2		1	0.95	**0.99**	0.95	0.93	0.52	**0.99**
Set 3			1	0.95	**0.99**	0.85	0.28*	0.95

* p-value > 0.05

The estimation of the correlation coefficient (r) through the application of a Mantel test embedded in GENALEX, showed different values depending on the marker type. Within each marker, two different matrix types, genetic distance and log (genetic distance) matrices were compared to the Euclidean distance and the log (Euclidean distance) matrices of environmental variables (climatic and bioclimatic) in order to find the highest correlation coefficients. The correlation coefficient for each marker was determined in comparison with the genetic distance and the log (genetic distance). The AFLP 12 *loci* dataset (abbreviated as AFLP-12) was tested separately from the AFLP original *loci* dataset (abbreviated as AFLP).

The highest estimated correlation coefficients obtained from the Mantel test between genetic distance of each marker and Euclidean distance and/or log (Euclidean distance) matrices of the environmental variables were 0.56 (p-value = 0.03) for AFLP-12 – set 3, 0.52 (p-value = 0.00) for ITS1 – Tmax, 0.51 (p-value = 0.00) for AFLP – Tmax, 0.43 (p-value = 0.01) for log (SSR) – Tmin and 0.38 (p-value = 0.01) for log (mtDNA) – set 3. The ITS2 and cpDNA markers did not score any correlation value below the 0.05 p-value (Table 3.36).

All correlations were significantly high (above 0.50) between Euclidean distance matrices based on environmental variables and genetic distance matrix generated for AFLP-12, AFLP and ITS1; SSR *loci* and mtDNA region showed significant moderate (below 0.50) values; the ITS2 and cpDNA regions showed the lowest over all correlations (Fig. 3.28). By using the log (genetic distance) matrix, the overall correlations increased for mtDNA and SSR regions remarkably, but both ITS2 and cpDNA regions showed no improvement in overall correlations (Fig. 3.29).

Finally, correlations were summarized only for the highest significant values found between genetic distance and log (genetic distance) of each marker and Euclidean distance matrices of the successfully correlated environmental variables (set 3, Tmax and Tmin). The correlation scores were situated between -0.20 to 0.60. ITS1 was highly correlated to Tmax, log (mtDNA) to set 3, log (SSR) to Tmin, AFLP to Tmax and AFLP-12 to set 3 (Fig. 3.30).

The 12 AFLP *loci* previously detected were found to be correlated with the bioclimatic belt through the statistical relation found with the precipitation of driest month, precipitation of driest quarter and precipitation of warmest quarter (set 3). The same case was found for the mtDNA region but with a lower correlation coefficient. However the AFLP original *loci* and ITS1 were found to be correlated with the average of maximum temperature values (Tmax) during five months assumed for samples life cycle. SSR markers were the only one correlated with the average of minimum temperature values (Tmin) during the same five months. ITS2 and cpDNA showed no significant correlation with other tested climatic and bioclimatic variables

To test the correlation between the genetic markers, a Mantel test was used. The highest significant correlation coefficient (p-value < 0.05) for ITS1 was 0.64 with mtDNA, for ITS2 was 0.64 with cpDNA and *vice versa*, for mtDNA it was 0.78 with AFLP-12, for SSR it was 0.66 with AFLP, and as expected, AFLP was highly correlated (0.71) with AFLP-12 (Table 3.37).

Table 3.36 Correlation coefficients and p-values of Mantel test between: genetic distance matrix and log (genetic distance) of each marker against Euclidean distance and log (Euclidean distance) matrices of the environmental variables. Significant correlation value are written in bold, while for each marker the highest and most significant correlation values are highlighted in black.

Euclidean matrix of environmental variables		Genetic distance matrix													
		ITS1	Log (ITS1)	ITS2	Log (ITS2)	cpDNA	Log (cpDNA)	mtDNA	Log (mtDNA)	SSR	Log (SSR)	AFLP	Log (AFLP)	AFLP-12	Log (AFLP-12)
Geographic distance (km)	Distance	-0.08	-0.10	-0.17	-0.18	-0.16	-0.13	-0.09	-0.14	0.16	0.16	0.13	0.18	-0.01	**-0.21***
	Log (distance)	-0.03	-0.09	-0.09	-0.09	-0.11	-0.12	-0.03	-0.09	0.13	0.16	0.14	0.12	0.10	-0.08
19 bioclimatic variables	Set 1	**0.33***	0.20	0.05	0.16	-0.07	-0.06	**0.22**	**0.24**	**0.33***	**0.40****	**0.42****	**0.40****	**0.48****	**0.13***
	Log (set 1)	0.25	0.12	0.08	0.16	-0.04	-0.06	0.21	0.22	**0.33***	**0.43***	**0.40**	**0.43****	**0.48****	0.10
	Set 2	**0.38****	0.25	0.10	0.22	-0.03	-0.01	**0.27**	**0.32***	0.25	**0.35***	**0.39***	**0.37****	**0.53****	0.19
	Log (set 2)	0.26	0.18	0.09	0.18	-0.01	0.03	0.22	**0.27***	**0.24***	**0.33***	**0.30***	0.35	**0.46****	0.11
	Set 3	0.43	0.3	0.15	0.29	0.00	0.02	**0.31**	**0.38****	0.22	0.33	**0.39***	**0.35****	**0.56****	**0.29***
	Log (set 3)	0.34	0.27	0.17	0.28	0.06	0.11	**0.29**	**0.37***	0.21	**0.32***	0.30	**0.33****	**0.50****	0.20
	Single	0.06	0.00	-0.12	-0.10	-0.12	-0.13	-0.01	-0.05	**0.33***	**0.31***	0.26	**0.29***	0.11	-0.17
	Log (single)	0.03	-0.08	-0.10	-0.11	-0.11	-0.17	0.02	-0.04	**0.35****	0.35	**0.31***	**0.40****	**0.18***	-0.14
Three climatic variables of five months [x]	Tmax	**0.52****	0.39	0.00	0.17	-0.09	-0.11	**0.31**	0.28	**0.34***	**0.40****	**0.51****	**0.39****	**0.50****	**0.19***
	Log (Tmax)	**0.38***	**0.31***	-0.01	0.16	-0.04	-0.04	0.26	**0.25****	**0.33****	**0.38****	**0.39****	**0.34***	**0.41****	0.12
	Tmin	**0.44****	**0.32***	-0.06	0.10	-0.12	-0.13	**0.25**	0.20	**0.39***	**0.43****	**0.50****	**0.41***	**0.43****	0.09
	Log (Tmin)	0.28	0.21	-0.10	0.05	-0.11	-0.10	0.18	0.14	**0.35****	**0.39****	**0.38****	**0.35****	**0.32****	0.05
	P	**0.49****	**0.36***	-0.03	0.14	-0.12	-0.13	0.27	0.24	**0.33***	**0.39****	**0.50****	**0.40****	**0.47****	0.14
	Log (P)	**0.35***	0.27	-0.04	0.11	-0.07	-0.07	**0.22**	0.19	**0.31***	**0.37****	**0.38****	**0.35****	**0.38****	0.06

[X] See table 2.8, pp. 88, * 0.05 > p-value > 0.01, ** p-value ≤ 0.01

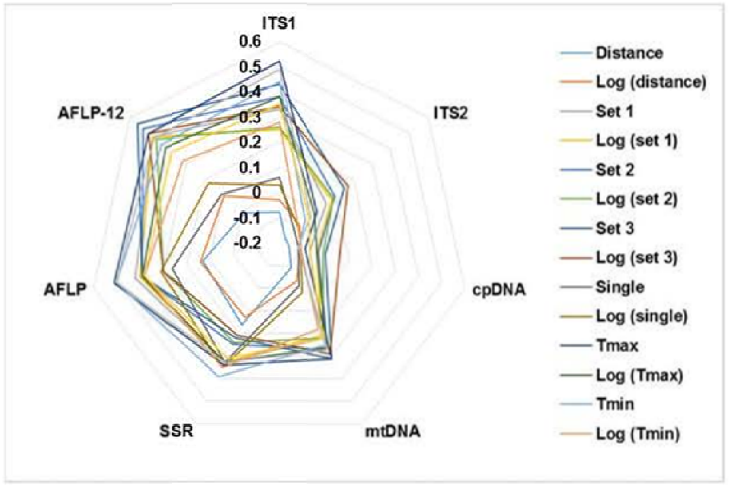

Figure 3.28 Radar chart for correlation coefficients values scored between genetic distance of each marker and Euclidean distance and/or log (Euclidean distance) matrices of the environmental variables. Correlation score is situated between -0.20 to 0.60. AFLP-12, AFLP and ITS1 regions show the highest overall correlations, followed by mtDNA and SSR. ITS2 and cpDNA show very low overall correlations.

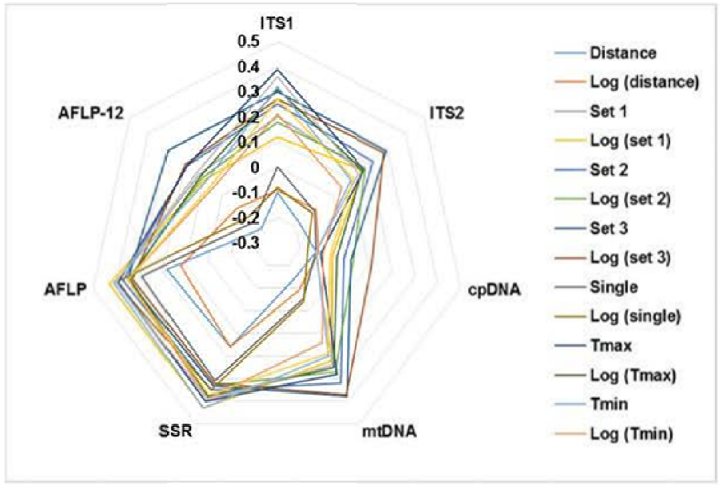

Figure 3.29 Radar chart for correlation coefficient values scored between log (genetic distance) of each marker and Euclidean and/or log (Euclidean) distance matrices of the environmental variables. Correlation score is situated between -0.30 to 0.50. SSR and mtDNA overall correlations show an improvement and become the highest correlation along with AFLP. ITS2 and cpDNA show no improvement in overall correlation to ecological variables.

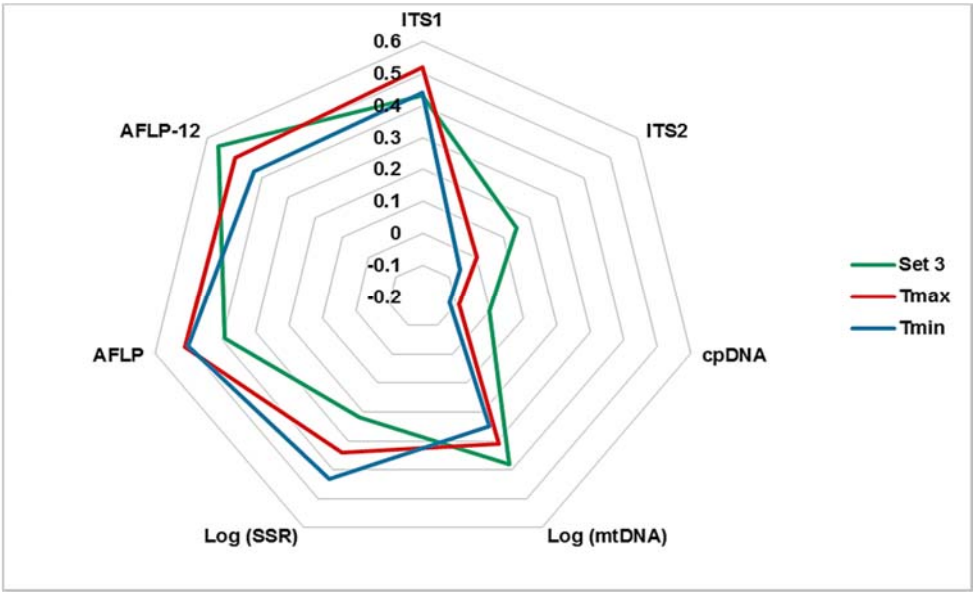

Figure 3.30 Radar chart for correlation coefficients values between genetic distance and log (genetic distance) of each marker and Euclidean distance matrices of successfully correlated environmental variables (set 3, Tmax and Tmin). Correlation scores are situated between -0.20 to 0.60. ITS1 was highly correlated to Tmax, log (mtDNA) to set 3, log (SSR) to Tmin, AFLP to Tmax and AFLP-12 to set 3.

Table 3.37 Correlation tests between all tested molecular markers. Highest values are written in bold.

Marker	ITS1	ITS2	cpDNA	mtDNA	SSR	AFLP	AFLP-12
ITS1		0.27	0.42	**0.64**	0.60	0.60	0.62
ITS2			**0.64**	0.35	0.27*	0.28*	0.51
cpDNA				0.46	0.40	0.26*	0.43
mtDNA					0.54	0.58	**0.78**
SSR						**0.66**	0.58
AFLP							**0.71**
AFLP-12							

***** p-value > 0.05

CHAPTER 4

DISCUSSION

4.1 METHODOLOGICAL ACHIEVEMENTS

This study planned to demonstrate the genetic variability of *Funaria hygrometrica* through DNA sequencing and DNA fingerprinting techniques in a highly diverse landscape. To achieve the first goal it was necessary to improve and/or develop new molecular markers. The second important goal was to find the relation between the detected genetic variability and the adaptation process to different climatic belts of the Sierra Nevada Mountains in Spain.

The DNA regions selected as candidates for sequencing were previously used in Funariaceae (Liu *et al*. 2012a), with the exception of the chloroplastic *rps*3-*rpl*16 region, a very interesting region for its variability, but so far underutilized (Borsch & Quandt 2009). The *rpl*16 intron is flanked by a very short exon at the 5'-region of the *rpl*16 gene and a major exon at the 3'-region. The *rpl*16 intron was used in previous phylogenetic and/or population studies in bryophytes, especially pleurocarpous species (e.g. Hedenäs & Eldenäs 2007; Huttunen *et al*. 2008; Hedenäs 2012a, b) but also some acrocarpous mosses like family Bryaceae (e.g. Pedersen & Hedenäs 2003; Pedersen *et al*. 2003). Initially the primers used with other bryophytes for the *rpl*16 intron were tried, but without success. Therefore new primers were designed with the aim to amplify the *rps*3-*rpl*16 intergenic spacer along with the *rpl*16 intron, which increase the acquired information. The primers were designed based on *Physcometrilla patens* data. The amplification and sequencing with the newly designed primers for this study were successful. Sequencing of this interesting region using our primers might be possible in other acrocarpous mosses.

Also some modifications to the ITS1 forward primer and the ITS2 reverse primer were introduced to improve its amplification quality, especially with some old herbarium specimens that could not be amplified using the ITS2 reverse primers designed by Douzery *et al*. (1999) or the ITS4 primer of White *et al.* (1990).

Mitochondrial *loci* are commonly much less targeted than chloroplast or nuclear *loci* for phylogenetic studies in bryophytes (Stech & Quandt 2010). In

Funaria hygrometrica, an unexpected high variability was found in *rpl*5-*rpl*16 mitochondrial DNA spacer. This shows that mitochondrial DNA sequences can be useful in population level studies of variable species.

In addition to sequencing nuclear, chloroplast and mitochondrial markers, DNA fingerprinting techniques were used to complement the analysis, for being more polymorphic and targeting more *loci* (Gee 2003). In this way, more information can be obtained for a better estimation of the genetic variability. Microsatellites are known to be polymorphic on species level and for their complicated form of evolution (Eisen 1999). Nevertheless, the degree of variability depends on the organism type, time since separation of lineages and the ploidy state of the organism (e.g. Goldstein *et al*. 1996). AFLP are known to amplify random regions extensively across the whole genome, increasing the probability to target regions involved in a biological processes (e.g. Wilding *et al*. 2001; Paris *et al*. 2010) or to gather as much information as possible to estimate the genetic variability based on data representing the whole genome (e.g. Fernandez *et al*. 2006). However, the targeted regions are random, and might consist of unnecessary junk DNA (i.e. stretches of DNA that do not code for genes), which could mislead the estimation of the real genetic variability (Bonin *et al*. 2006).

In this study, both sequencing and fingerprinting techniques were combined to benefit the advantage of each marker, to scan more genetic *loci* and to avoid any misleading while estimating the genetic variability and population structure.

4.2 GENETIC DIVERSITY

4.2.1 DNA polymorphism

Six different markers were used, four sequenced regions (ITS1 spacer, ITS2 spacer, *rps*3-*rpl*16 cpDNA region and *rpl*5-*rpl*16 mtDNA spacer) and two fingerprinting techniques (microsatellite and AFLP). DNA polymorphism and recombination rates were calculated in order to determine the genetic variability of each marker and its conservation level among and within locations.

4.2.1.1 Sequenced regions

Haplotypes were generated based on both polymorphic sites and indels. A total of 22, 7, 25 and 11 haplotypes were found for ITS1, ITS2, cpDNA and mtDNA *loci*, respectively. Haplotype diversity was 0.52, 0.37, 0.72 and 0.43 for ITS1, ITS2, cpDNA and mtDNA *loci*, respectively. Consequently, the chloroplast *rps*3-*rpl*16 region showed the highest haplotype diversity.

The nucleotide diversity was 0.63 %, 0.29 %, 0.21 % and 0.03 % for ITS1, ITS2, cpDNA and mtDNA *loci*, respectively. This reflects the conservation level of each *locus*, as cpDNA and mtDNA are more conserved (possess lower levels of nucleotide diversity) than ITS1 and ITS2.

4.2.1.2 Microsatellite

The results obtained by the microsatellite analysis of the five SSR *loci* showed differences in molecular weight apart from those shared by Dr. Goffinet (see Table 2.4, pp. 72). Unbiased gene diversity (UH_z) was 0.45, while the mean value of polymorphic index content (PIC) was 0.41. This reflects a moderate polymorphism level among sampled locations in comparison to the other markers.

Funaria hygrometrica as a haploid plant is expected to have only one copy per sample of the amplified SSR *locus*. However, in this study up to two

copies were detected with high band-density. One reason is due to the usage of a mixture a gametophytes obtained from one sporophyte (n = 2) as the source of DNA. This has the same effect as the utilization of sporophytic tissue, because the genetic diversity present in the sporophyte is represented. Another complication is the multiple chromosome numbers (x = 14, 28, 42 or 56) reported in *Funaria hygrometrica* (Fritsch 1982), which indicates the existence of different ploidy levels in this species. In mosses, polyploidy was previously deduced based on microsatellite patterns (e.g. Karlin *et al.* 2009).

Multiple alleles were detected in the SSR *loci* M3, M12 and M13 with 5, 7 and 3 alleles *per locus*, respectively (see Table 3.23, pp. 161). The present multi-allelic *loci* confirm a sufficiently high mutational rate of the amplified SSR *loci* to generate allelic variation.

4.2.1.3 AFLP

The AFLP technique permits a genome-wide scan of the genetic variability with a high number of variable markers. Therefore, there is a relative good chance to detect markers under selection either directly or because they are located near a gene under selection. The high reading output and the extensive statistical refining were expected to reflect more clearly the genetic variability of the studied samples.

AFLP marker data reported in bryophytes showed a low to moderate number of amplified bands and even modifications to the original protocol were conducted as adaptions to small genomes (e.g. mosses; Fernandez *et al.* 2006; Mikulášková *et al*. 2012). However, in *Funaria hygrometrica* an adequate number of bands were amplified with three selective bases at the EcoRI and MseI primers, which is the standard in genomes of average size.

A total of 3 057 bands were scored from all primer pairs for the 84 *Funaria hygrometrica* samples. Only 1 584 *loci* were adequate for further analysis, to avoid any misleading information due to statistical errors following the indications of Bonin *et al.* (2007) who recommends to discard all *loci* present below 5 % in the whole samples. The inbreeding coefficient F_{IS} was

adjusted to 1 as for a haploid, self-fertilizing plant, with no inbreeding depression (Taylor *et al.* 2007). The mean expected heterozygosity under Hardy-Weinberg assumption (H_e) was 0.21 which reflects the lowest detected diversity among the used genetic markers.

4.2.2 Neutrality state

The biological potential of a genotyped *locus* to reflect evolutionary events can be interpreted according to its neutrality state (Tajima 1989). A neutral state of a genetic region means that it evolves randomly, and is not affected by any non-random factor including selection, demographic expansion, genetic hitchhiking, introgression, etc. Neutral regions are not expected to be of any evolutionary importance in general. Chloroplast and mitochondrial markers behave in a special way as they are inherited without a recombination process comparable to the meiosis of the nuclear DNA. As a consequence, all regions are linked and any region of their genome, even without evolutionary importance, is affected when another region behaves in a non-neutral manner. When a neutral allele is favored by a linked gene under positive selection, the process is called genetic hitchhiking.

Tajima's neutrality test showed that ITS2 was neutral (although it has a negative D value), while ITS1, cpDNA and mtDNA were not neutral (significantly negative D value). A negative Tajima's D means an excess of low frequency polymorphisms relative to expectation, indicating population size expansion (e.g. after a bottleneck or a selective sweep) and/or purifying selection (Tajima 1989). However, in our case, more data are needed in order to find exactly the reason. At the end of this discussion a hypothetical scenario will be proposed to try to explain a presence of purifying selection under environmental stress.

4.2.3 Secondary structures

Haplotypes generated from the mitochondrial region *rpl*5-*rps*16 can be distinguished into two groups, one of which contained a fragment of 45 bp that was missing in the other group. By applying secondary structure prediction, we found that the 45 bp fragment forms a stem-loop structure (see Fig. 3.11, pp. 153). Stem-loops are believed to occur during single-stranding events when inverted repeats meet to form a region of pairing (the stem) surmounted by their interceding sequence (the loop; Kelchner 2000). In non-coding regions of organelle genomes stem-loop secondary structures are commonly reported (e.g. Michel *et al*. 1989; Sang *et al*. 1997; Downie *et al*. 1998). Loop regions of stem-loop secondary structures are often associated with hot spots for mutation in non-coding regions, affecting both nucleotide substitutions and indel events (Kelchner & Clark 1997; Kelchner 2000).

Although indels are most common in the terminal loops, they may occur anywhere along a secondary structure (Kelchner 2000). Occasionally, small segments of the stem itself will be deleted, decreasing the stem length, though perhaps not to an extent that would annihilate possible secondary structure formation (Kelchner 2000). Kelchner & Clark (1997) detected a deletion of a small sub-loop positioned partway up the stem of an *rpl*16 intron stem-loop in *Oryza sativa* L. Similarly, the secondary structure of *Funaria hygrometrica* based on sequences of the *rps*3-*rpl*16 cpDNA region contained an indel of 17 bp which forms part of a stem that is present in a haplotype group and absent in another (see Fig. 3.8, pp. 144).

Secondary structures have been widely discussed for ribosomal DNA, with ITS2 spacer, which is of special interest in plant systematics. The secondary structures of ITS2 and its phylogenetic implications were previously reported. For example, Baldwin *et* al. (1995) showed by the preliminary secondary structure analyses of ITS2 in *Calycadenia* DC. (Asteraceae) that there is no definite evidence of compensatory spacer mutation. Keller *et al.* (2008) found in the green algae, a branched helix I in the secondary structure of the ITS2 evolves past the '*Sphaeroplea*' clade. Their results corroborate the fundamental relevance of including the

secondary structure in sequence analysis and phylogenetics. In this work secondary structure prediction was used to improve the phylogenetic signal with the web-based toolkit (ITS2 database server), which uses the Neighbor Joining method to find the best tree. A program that would allow to combine both secondary structure information and Bayesian inference in an easy and straightforward way that could help and improve the phylogenetic signals.

4.2.4 High genetic diversity in Sierra Nevada Mountains

In spite of the relative small size of the sampling zone, an unexpected high genetic variability was found in *Funaria hygrometrica*. In a similar case, in Sierra Nevada, two distantly related genotype groups of the cosmopolitan moss *Ceratodon purpureus* (Hedw.) Brid. were reported based on cpDNA and nrDNA sequences (Mohamed Rizk 2012). Until now one of these two groups is only known from Sierra Nevada and Central Spain. The clear genetic distance between the two groups from Sierra Nevada exceeds the reported genetic distance between more geographically distant regions described before. Evidence for a possible natural hybridization between the different lineages of *Ceratodon purpureus* present in Sierra Nevada was found. This is in contrast with earlier reports that show that artificially produced hybrids between distant lineages within *Ceratodon purpureus* show a reduced viability. The results also reflect the high genetic variability at such a relatively small distance (Rizk 2012).

The cosmopolitan moss *Bryum argenteum* Hedw. is another moss studied in the Sierra Nevada Mountains, with a very high genetic diversity, but in *Bryum argenteum* the diversity is clearly structured according to altitude (Pisa *et al.* 2010). Based on nrITS and chloroplast *rps*4 sequence data, one clade included samples found in Sierra Nevada only above 2000 m a.s.l. This clade was known before only from the Antarctic region and might be restricted to cold environments (Pisa *et al.* 2010).

In contrast, recently studies on the widely distributed moss *Leucodon sciuroides* (Hedw.) Schwägr. revealed a relative low diversity between distant sampling locations (Stech *et al.* 2011; Werner & Ros unpublished data).

Samples were collected from North America, the Canary Islands, Cape Verde, several European countries, North Africa and several regions distributed over Russia. In this case, only three variable sites in ITS1 and four variable sites in ITS2 region were found. The estimated mean diversity overall samples was calculated based on ITS1 and ITS2 sequences of *Leucodon sciuroides* shared by the authors. The mean diversity of the samples of *Leucodon sciuroides* based on ITS1 and ITS2 was 0.05 % and 0.15 %, respectively. In *Funaria hygrometrica* the mean diversity over all samples calculated based on ITS1 and ITS2 from Sierra Nevada alone was 0.58 % and 0.30 %, respectively.

Werner *et al.* (2005) studied the rare moss *Pohlia bolanderi* (Lesq.) Broth. in Sierra Nevada Mountains of Spain by applying the ISSR technique (inter-simple sequence repeat). In general they found the genetic diversity within the species at one location in the range from 0.00 to 0.47. However, in *Funaria hygrometrica* the mean genetic diversity *per* location (analogue to heterozygosity) ranged from 0.23 to 0.60 for SSR and 0.13 to 0.44 for AFLP The difference may in part be due to the election of the ISSR genetic marker system that is evolutionary different from SSR and AFLP techniques. Another explication is that these authors centered their interest in a rare species. It is generally known that under the neutral theory rare species are less variable than frequent species (Nei 1987; Frankham 1996). Apart from the lower genetic variability mentioned before, the results were similar to those presented here in the sense that there is a certain differentiation between the different populations, but the major part of the variability in the case of *Funaria hygrometrica* corresponds to differences between individuals within populations and that not all individuals of one population cluster together.

Shaw *et al*. (2003) measured the quantitative morphological characters and sequenced the ITS region in New and Old World mosses (*Claopodium whippleanum* (Sull.) Renauld & Cardot*, Dicranoweisia cirrata* (Hedw.) Lindb. and *Scleropodium touretii* (Brid.) L.F. Koch) to find any subtle phenotypic differentiation between the disjunct populations through the Madrean-Tethyan distribution pattern (western North America and the Mediterranean region of southern Europe, northern Africa, and western Asia). Little or no

morphological differentiation existed between New and Old World plants in any of the species, but internal transcribed spacer (ITS) sequences revealed some phylogeographic structure. New World populations were more variable than Old World populations at the molecular level in the three species. Recent long-distance dispersal is a likely explanation for intercontinental disjunctions in these species. In our study, long-distance dispersal can be the explanation why *Funaria hygrometrica* haplotypes of the New World (U.S.A. accessions, see table 2.2, pp. 60) are closely related to overseas samples in Sierra Nevada Mountains of Spain.

Together, the cases of the common mosses (*Ceratodon purpureus*, *Bryum argenteum* and *Leucodon sciuroides*) and the overseas distribution pattern explained by long-distance dispersal, suggest that the extremely diverse landscape of Sierra Nevada of Spain accommodates a surprisingly high genetic diversity, which in some cases even passes the diversity found in prior studies for other bryophytes at a much larger, even worldwide scale. Sierra Nevada therefore not only possesses rare and endemic species but also an important part of hidden diversity in frequent and widely distributed species.

4.3 POPULATION STRUCTURE

4.3.1 Multiple introduction and colonization events

Based on STRUCTURE results (fingerprinting techniques only), samples that are genetically related are from different geographical locations, while samples from the same location do not always group together genetically. For example, based on microsatellite techniques (see Fig. 3.18, pp. 163), sample 8 (location 4, at 287 m a.s.l.), sample 51 (location 10, at 1 295 m a.s.l.) and sample 1 (location 1, at 24 m a.s.l.) are grouped together with others samples, although they belong to distant locations. In the AFLP technique (see Fig. 3.24, pp. 171), sample 19 (location 6, at 630 m a.s.l.), sample 62 (location 11, at 1 328 m a.s.l.), sample 13 (location 5, at 1 667 m a.s.l.) and sample 69 (location 12, at 1 650 m a.s.l.) had exactly the same similarity membership coefficient (i.e. a value in which a sample is assigned to a certain group) although they originate from distant locations.

Based on phylogenetic trees (sequenced *loci*), we can see the same case, as samples from different geographical locations group together. For example, samples 22, 23 and 24 (location 7, at 657 m a.s.l.), sample 37 (location 9, at 757 m a.s.l.) and samples 66 and 68 (location 12, at 1 650 m a.s.l.) were always clustered together with other samples of all sequenced *loci* (ITS1: see Fig. 3.2, pp. 126; ITS2: see Fig. 3.6, pp. 136; cpDNA: see Fig. 3.9, pp. 146; mtDNA: see Fig. 3.12, pp. 155).

Therefore, multiple introductions are inferred among sampled locations of *Funaria hygrometrica* in Sierra Nevada Mountains. In addition, *Funaria hygrometrica* is one of the few examples of the relatively rare fugitive life strategy species, which can occupy temporarily unexpected sites (During 1979, 1992; Frahm 2008). Multiple introduction and fugitive colonization strategy, both increase the ability of *Funaria hygrometrica* to persist under harsh conditions and adapt rapidly when it colonizes new locations (Dlugosch & Parker 2008).

4.3.2 Gene flow and genetic differentiation

The analysis of molecular variance (AMOVA) produces estimates of variance components and F-statistics analogs, reflecting the correlation of haplotypic diversity at different levels of hierarchical subdivision (Excoffier *et al*. 1992). In our case, AMOVA results were significant for all markers (p-value < 0.05). Gene flow (N_m) is a major factor impacting genetic structure and genetic differentiation among populations. The mean gene flow (N_m) estimated from F_{ST} and its analogue Φ_{ST} for ITS1, ITS2, cpDNA, mtDNA, SSR, AFLP and AFLP-12 were 0.85, 0.44, 0.52, 0.27, 2.00, 1.88 and 0.30 respectively. F_{ST} and its analogue Φ_{ST} were all above 0.25, which according to Wright (1978) indicates very great genetic differentiation. However, the detection of gene flow shows that the genetic differentiation among locations is not complete, in other words, a complete reproductive isolation between the sampled locations in Sierra Nevada Mountains is not the case. Wright (1931) proposed that when the gene flow among the locations $N_m > 1$, the homogenization is the result; when $N_m < 1$, the locations can be strongly differentiated. According to these criteria, strong genetic differentiation exists between the locations based on the sequenced regions, with the exception of SSR and AFLP markers which show some homogenization. However, the AFLP 12 *loci* data alone, showed the lowest gene flow value ($N_m = 0.30$), which reveal the highest genetic differentiation of all studied markers.

Loci with a significantly reduced level of spatial genetic differentiation are considered to be under balancing selection (e.g. Schierup 1998; Muirhead 2001; Willing *et al*. 2010). The detected AFLP 12 outlier *loci* show a high genetic differentiation level which is a signature of *loci* under selection. The genetic differentiation values for ITS1, ITS2, cpDNA and mtDNA are also relative high for an organism with a high dispersal capacity at relatively close distances. This might indicate that these *loci* are under positive selection or that there is no equilibrium between genetic drift and migration.

4.4 TWO GENETICALLY DIVERGENT LINES OR CRYPTIC SPECIES?

This question was raised after the constant grouping of the samples into two main groups based on every studied marker. Group A, represented constantly in blue in tables and figures, always includes the most abundant haplotypes and group B, represented constantly in red, always includes haplotypes less abundant than group A haplotypes and separated by certain genetic distance. However, the comparative scheme between all studied markers showed that some samples (2 to 4, 60, 63 to 65, 67, 70, 73, 76, 78 to 82 and 84) contain haplotypes of different groups when different markers are compared. These samples are referred to as non-homogenate samples.

Samples 60, 63, 70, 76, 80 and 81 of highland locations were found to be non-homogenate on nuclear DNA level (i.e. samples with haplotypes belonging to group A on ITS1 level and group B on ITS2 level or *vice versa*). In this case, combinations between the two groups on the nuclear level was found.

Samples 2 to 4, 73, 79, 81, 82 and 84 were found to be non-homogenate samples on maternal DNA level (i.e. samples with haplotypes belonging to group A on mtDNA level and group B on cpDNA level or *vice versa*). This is surprising because generally it is assumed that both chloroplasts and mitochondria are inherited exclusively by the maternal line in bryophytes (Natcheva & Cronberg 2007). Some researchers question the absolute case of the maternal inheritance of mitochondria and chloroplasts in plants (Terao 1918; Yang *et al*. 2000; Svab & Maliga 2007; McCauley *et al*. 2007). In *Funaria hygrometrica* a paternal source might cause the non-homogenate state on maternal DNA level. Such samples are indicator of a reproductive activity between the two groups, which rejects the possibility of a complete reproductive isolation between the two groups in the study area.

The analyses of linkage disequilibrium show clearly that there is a non-random association between genotypes of the different sequenced *loci*. Only locations 12 to 17 do not show linkage disequilibrium with the exception of the ITS1 and ITS2 *loci*, which are genetically linked by the proximity. A preferential association of independent *loci* can be the result of various factors

(Hartl & Clark 1997). Mutations are only important in regions without recombination as is the case in sex-chromosomes. Else recombination leads to a rapid decay of the disequilibrium. Random drift is only important in populations with a small relative population size and *loci* with a low recombination rate. This is the case of sex chromosomes or regions with chromosomal inversions. In our case, nuclear DNA on the one hand and, mitochondrial and chloroplast DNA on the other hand should be recombining freely and the population size is not low so that random drift is probably not involved. Inbreeding can lead to a considerable disequilibrium, essentially because it slows down the decay process. Although the degree of self-fertilization in *Funaria hygrometrica* is currently not known, in any case there remains the fact that disequilibrium occurs at lower altitudes (below 1 300 m a.s.l.) but not in the populations 12 to 17 situated at medium to high altitudes (above 1 300 m a.s.l.). This leaves the possibility of selection and population structuring in that highly diverse landscape.

Based on the fact that some crosses between the different lineages were detected, the hypothesis of having two cryptic species is rejected. The two major clades found in Sierra Nevada are closely related to haplotypes of different places worldwide as shown with the phylogenetic analysis of ITS1 and ITS2 regions. No clade was found to be exclusively restricted to Sierra Nevada Mountains. Fernandez *et al.* (2006) only used the AFLP *loci* to interpret the presence of cryptic species within the widely distributed moss *Grimmia laevigata* (Brid.) Brid. However, the usage of several independent markers based on both nuclear and extra-nuclear levels to study the presence of a cryptic species is preferable. In *Funaria hygrometrica*, AFLP would only reflect the presence of the two lineages as cryptic species. Combining AFLP data with mtDNA and/or cpDNA showed the fact that there is no reproductive isolation between the lines present in the sampled locations.

4.5 ADAPTATION TO BIOCLIMATIC BELTS

4.5.1 Outliers *loci*

Currently, there is an increasing interest in identifying genes or outlier *loci* that underlie adaptations to different factors in several species (e.g. Bonin *et al.* 2006; Nosil *et al.* 2008; Manel *et al.* 2010). Outlier *loci* are revealed by unusually high levels of population differentiation at specific marker *loci* (Black *et al.* 2001; Luikart *et al.* 2003; Storz 2005). Those *loci* that are involved in adaptation to local environmental conditions are indeed expected to exhibit increased differentiation among locations along with a decreased diversity within locations (Bonin *et al*. 2006).

For example, the study of the genetic frame of adaptation to a gradient of altitude in the common frog (*Rana temporaria* L.) by Bonin *et al*. (2006) showed that approximately 2 % of the AFLP *loci* they screened exhibited elevated altitudinal differentiation. In the case of spatially sympatric populations (i.e. separated populations that inhabit different environments) but exploit different ecological niches, it is possible to identify chromosomal regions involved in adaptive divergence by comparing relative levels of differentiation among multiple unlinked *loci* (Charlesworth *et al*. 1997; Stephan *et al*. 1998). Because local adaptation and directional selection should have *locus*-specific effects of reducing genetic variability within populations and increasing differentiation between populations, *loci* that are outliers for these characteristics are strong candidate regions for involvement in adaptation.

To the best of the present knowledge, this study is the first report on detection of candidate *loci* under selection by genome scan in the moss *Funaria hygrometrica* and might be the first report of all mosses. The AFLP genome scan analysis revealed 12 *loci* as under selection among a total of 1 584 *loci* scored in *Funaria hygrometrica*. These 12 *loci* possess a high credibility because, a) they were picked up by an exhaustive method (Dfdist embedded in MCHEZA software), b) a very stringent significance criterion of 99 % was applied, c) simulations were set up to the maximum number allowed by the program (100 000 simulation runs) and repeated at least 5 times, and d) a 30 % trimmed mean F_{ST} was applied to avoid spurious outlier

loci. Meyer *et al.* (2009) noted that the power of the analysis is directly associated with the genome coverage. Since a high number of reasonable size bands between 150 to 600 bp was found, the *loci* under selection detected here should prove to have a good reliability. The revealed 12 *loci* showed a variation continuum within all locations. They had an increased genetic differentiation among locations and decreased diversity within locations in comparison with the AFLP neutral *loci*. Considering all the previous points, the 12 outlier *loci* identified can be used for understanding the successful adaptation of *Funaria hygrometrica* through the different bioclimatic belts within Sierra Nevada Mountains of Spain.

4.5.2 Correlation tests

To elucidate the genetic bases of adaptation to different environments represents a goal of central importance and interest in evolutionary biology (Storz 2005). Magdy (2010) carried out a phylogenetic analysis for *Funaria hygrometrica* sampled from Sierra Nevada based on SRAP (Sequence-Related Amplified Polymorphism). A genetic structure related to altitudinal differences between locations of the lowland and midland up to 1 295 m a.s.l. was found. This led to suggest that a part of the genetic variability may be due to adaption to the different climatic conditions in the zone of study.

However, in this study, with a wider sampling range and through the phylogenetic analysis of the studied samples, it was not possible to find a specific clustering that reflects an altitudinal gradient (24 m up to 2 700 m a.s.l.). Therefore, the genetic distance based on genotyped regions and/or fingerprinting techniques were tested to the hypothesis of being statistically correlated to certain bioclimatic, climatic and/or geographical variables.

4.5.2.1 Correlation to maximum temperature gradient

During the assumed life cycle of the sampled *Funaria hygrometrica* from Sierra Nevada, the estimated average of maximum temperature was correlated to the ITS1 and the AFLP all *loci* dataset.

In the case of ITS1, as a non-coding region, it showed the ability to reflect indirectly the potential of the used samples to adapt along different maximum temperature ranges found between locations. One explanation, the ITS1 region might be linked to a near region under selection. However, ITS1 and ITS2 form a linkage group as they both belong to the nuclear ribosomal DNA (rDNA) unit 18S-ITS1-5.8S-ITS2-25S (Wheeler & Honycutt 1988). In our case, ITS2 was weakly related to ITS1 (r-value of 0.21), and did not reflect any adaptive value, though we did amplify the ITS 1 and 2 together (complete ITS region: ITS1-5.8S-ITS2), but ITS1 and 2 were sequenced separately using the modified primers (see Table 2.3, pp. 68). This redirects our suggestion that the ITS1 region is evolutionary parallel with an adaptive gene other being linked to such hypothetical genes. Buckler *et al*. (1997) demonstrated that rDNA repeats evolve together through concerted evolution. However, some genomes contain a considerable diversity of paralogue rDNAs (two copies diverged after a duplication event located at different sites in the genome). ITS1 undergoes a splicing process apart from ITS2, as the latter conserves a secondary structure form to complete the formation of the ribosomal small subunit (Keller *et al.* 2010). Therefore, it is possibly that the ITS1 and 2 would evolve separately which might cause the differences in its relation to adaptive process.

In the case of AFLP, the data give an overview over a wide part of the genome. Amplified bands might or might not correspond to functional regions, but can be linked to nearby genes under selection and therefore reflect indirectly the adaptation to the maximum temperature averages. More analyses will be proposed afterwards to ensure the nature of the amplified regions and their ability to adapt to the gradient of the maximum temperature averages.

4.5.2.2 Correlation to minimum temperature gradient

The estimated average of minimum temperature during *Funaria hygrometrica* life cycle in sampled locations was correlated to all amplified SSR *loci*. The tested SSR *loci* were able to reflect indirectly the potential of the samples to adapt along different averages of minimum temperature found between locations (see Fig. 3.29, pp. 182). SSR were proven to be distributed non-randomly across protein-coding regions, untranslated regions (UTRs), and introns (Kantety *et al*. 2002). SSR expansions and/or contractions in protein-coding regions can lead to a gain or loss of gene function via frame shift mutation or expanded toxic mRNA (Li *et al*. 2004). However, more studies are needed to determine if a certain SSR *locus* is located within a functional region or is linked to a specific gene that might have an influence in the adaptation process of *Funaria hygrometrica* to the differences in the gradient of the minimum temperature averages in Sierra Nevada Mountains.

Multiple alleles found in SSR analysis suggest that possibly the polyploidy might be a factor that controls the separation of the two lineages, or it would be of an adaptive value. Košnar *et al.* (2012) studied the role of polyploidy in the origin of European taxa of the moss *Tortula muralis* Hedw. No phylogenetic signal was associated with ploidy-level variation, suggesting a polytopic origin of the diploids samples. In our case, the role of polyploidy would be the same as proposed for *Tortula muralis*, but still more studies are needed in order to determine the exact polyploidy level in *Funaria hygrometrica* in relation to sequence markers and the role of ploidy levels in relation to adaption to environmental factors.

4.5.2.3 Correlation to the gradient of the temperature and precipitation

In a comparison with the other markers, the detected 12 AFLP outlier *loci* and the mtDNA sequence data were found to be highly correlated to the precipitation of driest month, precipitation of driest quarter and precipitation of warmest quarter. In detail, these variables represent the interaction between the drought periods and warm temperature during the year (Hijmans *et al*.

2005). However, the 12 AFLP *loci* were more correlated to such factors than the mtDNA (see Table 3.36, pp. 181).

The mtDNA region sequenced here, as a non-coding region, showed the ability to reflect indirectly the potential of the samples to be adapt along the gradient of temperature and precipitation interactions found between locations in Sierra Nevada (see Fig. 3.28, pp. 182). Even though this intergenic spacer is not a functional gene, the mitochondrial DNA is inherited as one linkage group in the absence of recombination (Hayashi *et al.* 1985, Hoech *et al.* 1991). Therefore, any hypothetical mitochondrial gene under selection would directly affect the frequency of sequence variants of all other regions of the mitochondrial DNA. Further studies are needed to identify the hypothetical mitochondrial genes participating in the adaptive process of *Funaria hygrometrica* in its habitats of Sierra Nevada.

As the detected AFLP *loci* are likely located in non-coding DNA, some of the outlier *loci* may only exhibit the signature of selection because they are linked to the actual target (Schlötterer 2003; Tollenaere *et al*. 2011). Although it is difficult to know the location and function of the *loci* involved in the adaptation to climatic factors, a genome scan of *Funaria hygrometrica* still offers a unique opportunity to unravel the genetic basis of moss adaptation without known phenotypes and whole genome sequences. In particular, the AFLP primers that were used to amplify the identified outlier *loci* can be directly used to construct a reduced representation library of the *Funaria hygrometrica* genome, which would allow efficient sequencing of the linked genomic regions by next-generation sequencing technology (Hohenlohe *et al*. 2012).

Finally, through the correlation test based on the genetic distance generated from six molecular markers, it was proved that *Funaria hygrometrica* possesses the ability to adapt genetically to the spatial gradient of the averages of maximum temperature, minimum temperature and the interaction between temperature and precipitation. Consequently, *Funaria hygrometrica* would be able to adapt to the climate changes may occur on the sampling locations in the future.

4.6 HYPOTHETICAL SCENARIOS

4.6.1 Observations

The high genetic variability together with the genetic differentiation between the sampled locations taken along an altitudinal gradient in Sierra Nevada of Spain spanning different bioclimatic belts raises the possibility of low migration. A low detected migration rate does not necessarily mean that the dispersal is low, it can also be explained by a low rate of establishment of spores when they encounter unsuitable environmental conditions for their genotype.

As an interesting observation, the non-homogenate samples are mostly present above 1 300 m a.s.l., except samples 2 to 4 from location 2 (58 m a.s.l.). Additionally, homogenate samples of the most abundant haplotypes (group A) found to be decreased in number above 1 300 m a.s.l., in contrast to homogenate individuals of the less abundant haplotypes (group B) and non-homogenate samples.

Another interesting observation is that no recombination was detected between the two lines below 1 300 m a.s.l., even between proximate locations (e.g. locations 6 and 7), despite of the high dispersal ability reported in *Funaria hygrometrica* (During 1979).

The lowland locations 2 and 7 with clearly distinct haplotype patterns are different from other lowland populations by the water availability. Location 2 was situated in a palm tree plantation with abundant artificial watering and location 7 was located at an especially moist soil near a water source. Sample 37 appeared to be the only sample that belongs to group B haplotypes in location 9 where the rest belong to group A haplotypes.

Based on such observations we suggest that, a) some factors above 1 300 m a.s.l. favor the crossing between the two lines on both maternal and nuclear DNA levels or, b) some factors limit the recombination between the two lines below 1 300 m a.s.l. This limitation might be related to the relatively high temperature of the lowlands and/or the drought of sampled locations which might limit the growth of certain genotypes.

4.6.2 Hypotheses

Migration levels of the studied regions suggest the presence of certain limitation to gene flow that might be due to environmental or genetic barriers. To analyze the actual barriers to gene flow a higher sample number and wider sampling range are recommended. On the other side, the detected level of gene flow might be due to human factors (e.g. Wen & Hsiao 2001) or historical migrations associated with past climate change (Ohsawa *et al*. 2007).

The human factor may play certain role as Sierra Nevada is one of the attractive mountains to be visited by tourists in Southern Europe. Tourists from different places all around the world aim the mountain for landscape visiting and skiing. Tours usually starting from the sea side of Motril up to Granada ending in ski-station at the top of the mountains which coincidence with the sampling route. Additionally, *Funaria hygrometrica* is known for following fugitive life strategy (During 1992) and is often found at burnt sites on road sides, where human presence is highly probable for camping and different activities. The presence of human beings in the studied region in evolutionary recent times could have augmented the number of suitable habitats for this species, e.g. along roadsides and this way have caused a population expansion. Population expansion could explain at least partly the significant negative values for Tajima´s D. As a consequence, the high genetic diversity found in Sierra Nevada might simply be the result of a high migration rate into a landscape formed by human factors. If this is the case, a uniform landscape of similar size and under similar human impact but with uniform environmental conditions should have a high genetic diversity as well. This hypothesis can easily be tested.

The hypothesis that the genetic diversity is at least in part due to selective advantages of certain lineages under specific environmental conditions would be confirmed if the genetic diversity in a uniform landscape is significantly lower than in Sierra Nevada. One result that favors this hypothesis is that Mantel tests found a significant correlation between the detected genetic diversity of the sampled locations and the averages of

maximum, minimum temperatures and the interaction between the temperature and precipitation. Significant negative Tajima´s D value of the ITS1 and mtDNA can also be explained by purifying selection due to environmental variables. The establishment of evolutionary favored lineages in a diverse landscape can be reached in two ways:

I. The selection of favored lineages can occur locally. Under this scenario an ancestral population present in Sierra Nevada developed *in situ* adaption to the different conditions present in the studied area. The local selection possibly was followed by migration under the effect of past environmental changes along the altitudinal gradient of the mountain. In any case it seems that there is some dispersal at the scale of the studied area, as the different lineages occur at all altitudes but at different frequencies.
II. Alternatively, another scenario is that the genetic variability could be the consequence of a constant migration of samples representing the genetic diversity of *Funaria hygrometrica* at wider scale. The dispersed spores will only germinate and give rise to viable populations, where suitable conditions for their survival are present and consequently the population will be able to reproduce. Under this scenario, the diverse landscape of Sierra Nevada would permit the establishment of genetically diverse lineages also present at many other sites, perhaps even at a worldwide scale.

In both ways, the environmental conditions of the locations are critical for selection, represented in the sampling site's temperature and precipitation. Both factors influence the availability of water in air and thus, would generate a differential factor for one of the haplotype group rather than the other. For example, the condition in which water is mostly available possibly permit samples 2 to 4 (location 2), 22 to 24 (location 7) and 37 (location 9) to survive and reproduce. The latter in which, microclimate effect would explain the presence of this genotype in a location where its samples belong to group A, except to sample 37 (Busby *et al*. 1978). However the apparent differences of locations 2 and 7 from the other lowland locations, could be that both detected lineages of *Funaria hygrometrica*, one is adapted especially to dry Mediterranean habitats. At higher altitudes the two lineages grow together

and cross, while at lower altitudes the supposed Mediterranean type is much more frequent, unless special environmental conditions make high amounts of water available. This hypothesis could be tested in the future comparing populations of *Funaria hygrometrica* from other regions with a pronounced Mediterranean climate but with different amounts of water availability.

It is possible to decide which of the two adaptive scenarios is preferable by comparing the variability of Sierra Nevada with the variability found at a larger scale. If local adaption is an important factor, the samples from Sierra Nevada should be clearly different from the samples found in other regions, especially if they are situated at large distances. But if the genetic variability is due to the establishment of migrants, similar or even identical genotypes should be found elsewhere, even at large distances, when they are sampled at locations with similar ecological conditions. At present the amount of sequence information from other sites is still limited, but the available ITS sequence data seem to point more in the direction of a high amount of long distance migration.

The explanation of the high genetic diversity at high elevation and crosses between the two found divergent lineages of *Funaria hygrometrica* is that the severe conditions at high altitudes trigger adaptation (Porter & Rice 2013). Seasonal variation of the cold tolerance of bryophytes suggests that mosses develop tolerance in response to changes in environmental conditions (Rütten & Santarius 1992, 1993). This raises an alternative hypothesis, that the genetic variation found in *Funaria hygrometrica* has no adaptive value, but that the high dispersal rate of this species leads to a high genetic diversity at a local scale in the sense of the “everything is everywhere” hypothesis discussed for many microorganisms (Fontaneto 2011).

However, the presented results favor especially the scenario that, the genetic variability is due to a constant migration of samples representing the genetic diversity of *Funaria hygrometrica* worldwide. And also support the theory that mosses achieve wide geographic distributions across large ecological gradients (e.g. mountains) through adaptive genetic variation (e.g. Korpelainen *et al*. 2012).

CHAPTER 5

CONCLUSIONS

Conclusions obtained from the current study can be summarized in the following points:

1- Successful new designed and/or modified primers were developed to improve the PCR amplification for the chloroplast *rps*3-*rpl*16 region and the nuclear ITS spacer, which are useful for population genetics and phylogenetic studies.

2- High genetic variability was found in the moss *Funaria hygrometrica* sampled from locations along an altitudinal gradient of Sierra Nevada Mountains of Spain in spite of the relative small size of the sampling area.

3- Strong genetic differentiation exists between the locations based on the four sequenced regions (ITS1 spacer, ITS2 spacer, *rps*3-*rpl*16 cpDNA region and *rpl*5-*rpl*16 mtDNA spacer) and two fingerprinting techniques (microsatellite and AFLP). However, the microsatellite and AFLP genetic markers, showed less differentiation. Nevertheless, the detected gene flow indicate that there are migrants between the studied locations and therefore are not completely isolated in reproductive terms.

4- The high genetic variability found within locations indicates that these locations were colonized independently several times. Multiple introductions and the fugitive life strategy followed by *Funaria hygrometrica* can efficiently accelerate the speed of adaptive evolution.

5- The analysis of the results obtained from the six genetic markers revealed the existence of two genetically divergent lineages of *Funaria hygrometrica* located in Sierra Nevada of Spain. The existence of samples with markers belonging to different lineages indicates genetic contact. A treatment of the two lineages as “cryptic” species is therefore not supported.

6- Correlation tests based on the genetic distance generated from the six molecular markers with spatial gradient of the averages of maximum temperature, minimum temperature, and the interaction between temperature and precipitation, indicate the possibility of adaptive genetic variation. Consequently, *Funaria hygrometrica*

would be able to adapt to the climate changes that may occur on the sampling locations in the future.

7- From the two detected lineages of *Funaria hygrometrica*, one might be adapted especially to dry Mediterranean habitats. At higher altitudes the two lineages grow together and cross, while at lower altitudes the supposed Mediterranean type is much more frequent, unless special environmental conditions make high amounts of water available.

8- The results favor the scenario that the genetic variability is due to a constant migration of samples representing the genetic diversity of *Funaria hygrometrica* at a wider scale.

9- The current results support the hypothesis that mosses achieve wide geographic distributions across large ecological gradients (e.g. mountains) through adaptive genetic variation.

CHAPTER 6

REFERENCES

Agapow P.-M. & Burt A. 2001. Indices of multilocus linkage disequilibrium. *Molecular Ecology Notes* 1: 101-102.

Ah–Peng C., Chuah–Petiot M., Descamps–Julien B., Bardat J., Stamenoff P. & Strasberg D. 2007. Bryophyte diversity and distribution along an altitudinal gradient on lava flow in La Réunion. *Diversity and Distributions* 13: 654-662.

Aitken S. N., Yeaman S., Holliday J. A., Wang T. & Curtis-McLane S. 2008. Adaptation, migration or extirpation: Climate change outcomes for tree populations. *Evolutionary Applications* 1: 95-111.

Altenberg L. 1995. Genome growth and the evolution of the genotype-phenotype map. *Lecture Notes in Computer Science* 899: 205-259.

Álvarez I. & Wendel J. F. 2003. Ribosomal ITS sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution* 29: 417-434.

Amarillo-Suarez A. R. & Fox C. W. 2006. Population differences in host use by a seed beetle: local adaptation, phenotypic plasticity and maternal effects. *Oecologia* 150: 247-258.

Anderson M. J. 2003. *PCO: a FORTRAN computer program for principal coordinate analysis*. Department of Statistics, University of Auckland, New Zealand.

Antao T. & Beaumont M. A. 2011. Mcheza: a workbench to detect selection using dominant markers. *Bioinformatics* 27: 1717-1718.

Armour J. A. L., Alegre S. A., Miles S., Williams L. J. & Badge R. M. 1999. Minisatellites and mutation processes in tandemly repetitive DNA. In: Goldstein D. B. & Schlotterer C. (eds.), *Microsatellites: evolution and applications*. Oxford University Press, Oxford, pp. 24-33.

Arrigo N., Tuszynski J. W., Ehrich D., Gerdes T. & Alvarez N. 2009. Evaluating the impact of scoring parameters on the structure of intra-specific genetic variation using RawGeno, an R package for automating AFLP scoring. *BioMed Central Bioinformatics* 10: 33.

Arrigo N., Holderegger R. & Alvarez N. 2012. Automated scoring of AFLPs using RawGeno V2, a free R CRAN library. In: Pompanon F. & Bonin A. (eds.), *Data production and analysis in population genomics: methods and protocols*. Methods in Molecular Biology series 888. Humana Press, New York, pp. 155-175.

Bailey C. D., Carr T. G., Harris S. A. & Hughes C. E. 2003. Characterization of angiosperm nrDNA polymorphism, paralogy, and pseudogenes. *Molecular Phylogenetics and Evolution* 29: 435-455.

Baird N. A., Etter P. D., Atwood T. S., Currey M. C., Shiver A. L., Lewis Z. A., Selker E. U., Cresko W. A. & Johnson E. A. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE* 3: e3376.

Baldwin B. G., Sanderson M. J., Porter J. M., Wojciechowski M. F., Campbell C. S. & Donoghue M. J. 1995. The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. *Annals of the Missouri Botanical Garden* 82: 247-277.

Balkenhol N., Waits L. P. & Dezzani R. J. 2009. Statistical approaches in landscape genetics: an evaluation of methods for linking landscape and genetic data. *Ecography* 32: 818-830.

Beckert S., Steinhauser S., Muhle H. & Knoop V. 1999. A molecular phylogeny of bryophytes based on sequences of the mitochondrial *nad*5 gene. *Plant Systematics and Evolution* 218: 179-192.

Birky C. W. Jr. 1995. Uniparental inheritance of mitochondrial and chloroplast genes: mechanism and evolution. *Proceedings of the National Academy of Sciences of the United States of America* 92: 11331-11338.

Black W. C. T., Baer C. F., Antolin M. F. & DuTeau N. M. 2001. Population genomics: genome-wide sampling of insect populations. *Annual Review of Entomology* 46: 441-469.

Blanca G. 1991. *Joyas botánicas de Sierra Nevada*. Editorial La Madraza. Granada.

Blanca G., Cabezudo B., Cueto M., Fernández López C. & Morales Torres C. 2009. *Flora Vascular de Andalucía Oriental*. *Volumen 1: Selaginellaceae–Ceratophyllaceae*. Consejería de Medio Ambiente, Junta de Andalucía, Sevilla.

Bohonak A. J. 2002. IBD (isolation by distance): a program for analyses of isolation by distance. *Journal of Heredity* 93: 153-154.

Bonin A., Bellemain E., Eidesen P. B., Pompanon F., Brochmann C. & Taberlet P. 2004. How to track and assess genotyping errors in population genetics studies. *Molecular Ecology* 13: 3261-3273.

Bonin A., Pompanon F. & Taberlet P. 2005. Use of amplified fragment length polymorphism (AFLP) markers in surveys of vertebrate diversity. *Molecular Ecology* 395: 145-161.

Bonin A., Taberlet P., Miaud C. & Pompanon F. 2006. Explorative genome scan to detect candidate loci for adaptation along a gradient of altitude in the common frog (*Rana temporaria*). *Molecular Biology and Evolution* 23: 773-783.

Bonin A., Ehrich D., Manel S. 2007. Statistical analysis of amplified fragment length polymorphism data: a toolbox for molecular ecologists and evolutionist. *Molecular Ecology* 16: 3737-3758.

Bopp M. 1961. Morphogenese der Laubmoose. *Biology Review* 36: 237-280.

Bopp M. 1976. External and internal regulation of the differentiation of the moss protonema. *The Journal of the Hattori Botanical Laboratory* 41: 167-177.

Bopp M. 1980. The hormonal regulation of morphogenesis in mosses. In: Skoog F. (ed.), *Plant Growth Substances*. Springer-Verlag, Berlin, pp. 351-361.

Borsch T., Hilu W., Quandt D., Wilde V., Neinhuis C. & Barthlott W. 2003. Non-coding plastid *trn*T–*trn*F sequences reveal a well resolved phylogeny of basal angiosperms. *Journal of Evolutionary Biology* 16: 558-576.

Borsch T., Löhne C., Müller K., Hilu K. W., Wanke S., Worberg A., Barthlott W., Neinhuis C. & Quandt D. 2005. Towards understanding basal angiosperm diversification: recent insights using rapidly evolving genomic regions. *Nova Acta Leopoldina* 92: 85-110.

Borsch T., Hilu K. W., Wiersema J. H., Löhne C., Barthlott W. & Wilde V. 2007. Phylogeny of *Nymphaea* (Nymphaeaceae): evidence from substitutions and microstructural changes in the chloroplast *trn*T-*trn*F region. *International Journal of Plant Science* 168: 639-671.

Borsch T. & Quandt D. 2009. Mutational dynamics and phylogenetic utility of noncoding chloroplast DNA. *Plant Systematics and Evolution* 282: 169-199.

Bowcock A. M., Linares A. R., Tomfohrde J., Minch E., Kidd J. R. & Cavalli-Sforza L. L. 1994. High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* 368: 455-457.

Brown W. M., George M. Jr. & Wilson A. C. 1979. Rapid evolution of animal mitochondrial DNA. *Proceedings of the National Academy of Sciences of the United States of America* 76: 1967-1971.

Brugués M. & Ruiz E. 2010. *Funaria*. In: Guerra J., Brugués M., Cano M. J. & Cros R. M. (eds.), *Flora Briofítica Ibérica, Volumen IV, Funariales, Splachnales, Schistostegales, Bryales, Timmiales.* Universidad de Murcia, Sociedad Española de Briología, Murcia, pp. 59-63.

Buckler E. S., Ippolito A. & Holtsford T. P. 1997. The evolution of ribosomal DNA divergent paralogues and phylogenetic implications. *Genetics* 145: 821-832.

Busby J. R., Bliss L. C. & Hamilton C. D. 1978. Microclimate control of growth rates and habitats of the boreal forest mosses, *Tomenthypnum nitens* and *Hylocomium splendens*. *Ecological Monographs* 48: 95-110.

Calonje M., Martín-Bravo S., Dobeš C., Gong W., Jordon-Thaden I., Kiefer C., Kiefer M., Paule J., Schmickl R. & Koch M. A. 2009. Non-coding nuclear DNA markers in phylogenetic reconstruction. *Plant Systematics and Evolution* 282: 257-280.

Campbell A. K. 2003. Save those molecules: molecular biodiversity and life. *Journal of Applied Ecology* 40: 193-203.

Campbell D. & Bernatchez L. 2004. Generic scan using AFLP markers as a means to assess the role of directional selection in the divergence of sympatric whitefish ecotypes. *Molecular Biology and Evolution* 21: 945-956.

Capesius I. 1995. A molecular phylogeny of bryophytes based on the nuclear encoded 18s rRNA genes. *Journal of Plant Physiology* 146: 59-63.

Capesius I. & van de Peer Y. 1997. The first secondary structure of the large ribosomal subunit RNA gene from the moss *Funaria hygrometrica*. *Journal of Plant Physiology* 151: 239-241.

Casgrain P. & Legendre P. 2001. *The R Package for multivariate and spatial analysis, version 4.0* - User's manual. Departement de Sciences Biologiques, Université de Montreal.

Cech T. R. 1998. Conserved sequences and structures of group I introns: building an active site for RNA catalysis – a review. *Gene* 73: 259-271.

Cech T. R., Damberger S. H., Gutell R. R. 1994. Representation of the secondary and tertiary structure of group I introns. *Journal of Structural Biology* 1: 273-280.

Chambers G. K. & MacAvoy E. S. 2000. Microsatellites: consensus and controversy. *Comparative Biochemistry and Physiology (Part B)* 126: 455-476.

Channell J. E. T. & Medizza F. 1981. Upper Cretaceous and Paleogene magnetic stratigraphy and biostratigraphy from the Venetian, Southern Alps. *Earth and Planetary Science Letters* 55: 419-432.

Chantanaorrapint S. 2010. *Ecological studies of epiphytic bryophytes along altitudinal gradients in Southern Thailand*. Universität Bonn, Ph.D. thesis.

Charlesworth B., Nordborg M. & Charlesworth, D. 1997. The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genetics Research* 70: 155-174.

Chien G., Crosby M. & He S. 2003. *Moss Flora of China*, *English Version, Volume 3: Grimmiaceae through Tetraphidaceae.* Science Press and Missouri Botanical Garden, Beijing, New York, St. Louis.

Chopra R. N. & Rashid A. 1967. Apogamy in *Funaria hygrometrica* Hedw. *The Bryologist* 70: 206-208.

Christianson M. L. & Hornbuckle J. S. 1999. Phenylurea cytokinins assayed for induction of shoot buds in the moss *Funaria hygrometrica*. *American Journal of Botany* 86: 1645-1648.

Christianson M. L. & Duffy S. H. 2002. Dose-dependent effect of salicylates in a moss, *Funaria hygrometrica*. *Journal of Plant Growth Regulation* 21: 200-208.

Churchill S. P., Griffin III D. & Muñoz J. 2000. A checklist of the mosses of the tropical Andean countries. *Ruizia* 17: 1-203.

Clarke L. J., Ayre D. J. & Robinson S. A. 2009. Genetic structure of East Antarctic populations of the moss *Ceratodon purpureus. Antarctic Science* 21: 51-58.

Colegrave N. & Collins S. 2008. Experimental evolution: experimental evolution and evolvability. *Heredity* 100: 464-470.

Coleman A. W. 2003. ITS2 is a double-edged tool for eukaryote evolutionary comparisons. *Trends in Genetics* 19: 370-375.

Coleman A. W. & Mai J. C. 1997. Ribosomal DNA ITS 1 and ITS 2 sequence comparisons as a tool for predicting genetic relatedness. *Journal of Molecular Evolution* 45: 168-177.

Conord C., Lempérière G., Taberlet P. & Després L. 2006. Genetic structure of the forst pest *Hylobius abietis* on conifer plantations at different spatial scales in Europe. *Heredity* 97: 46-55.

Cooke G., Chao N. & Beheregaray L. 2012. Divergent natural selection with gene flow along major environmental gradients in Amazonia: insights from

genome scans, population genetics and phylogeography of the characin fish *Triportheus albus*. *Molecular Ecology* 21: 2410-2427.

Coop G., Witonsky D., Di Rienzo A. & Pritchard J. K. 2010. Using environmental correlations to identify loci underlying local adaptation. *Genetics* 185: 1411-1423.

Cove D. 2000. The generation and modification of cell polarity. *Journal of Experimental Botany* 51: 831-838.

Cove D. 2005. The moss *Physcomitrella patens*. *Annual Review of Genetics* 39: 339-358.

Cox C. J., Goffinet B., Shaw A. J. & Boles S. B. 2004. Phylogenetic relationships among the mosses based on heterogeneous Bayesian analysis of multiple genes from multiple genomic compartments. *Systematic Botany* 29: 234-250.

Cronberg N., Natcheva R. & Hedlund K. 2006. Microarthropods mediate sperm transfer in mosses. *Science* 313: 1255.

Crosby M. R., Magill R. E., Allen B. & He S. 1999. *Checklist of the Mosses*. Missouri Botanical Garden, St. Louis.

Crow J. F. 1999. Hardy, Weinberg and language impediments. *Genetics* 152: 821-825.

Crum H. A. 1972. The geographic origins of the mosses of North America's eastern deciduous forest. *Journal of the Hattori Botanical Laboratory* 25: 269-298.

Crum H. A. 2001. Miscellaneous notes on *Sphagnum*-11. *Contributions from the University of Michigan Herbarium* 23: 107-114.

Darty K., Denise A. & Ponty Y. 2009. VARNA: interactive drawing and editing of the RNA secondary structure. *Bioinformatics* 25: 1974-1975.

Davis M. B. & Shaw R. G. 2001. Range shifts and adaptive responses to quaternary climate change. *Science* 262: 673-679.

Davis M. B., Shaw R. G. & Etterson J. R. 2005. Evolutionary responses to changing climate. *Ecology* 86: 1704-1714.

Davis M. L. & Cornwell D. A. 2006. *Introduction to Environmental Engineering*, 4th Edition. McGraw-Hill Education, India.

Deka R., Jin L., Shriver M. D., Yu L. M., DeCroo S., Hundrieser J., Bunker C. H., Ferrell R. E. & Chakraborty R. 1995. Population genetics of dinucleotide (dC-dA)n(dG-dT) polymorphisms in world populations. *The American Journal of Human Genetics* 56: 461-474.

Delaneau O. & Zagury J.-F. 2012. Haplotype inference. In: Pompanon F. & Bonin A. (eds.), *Data production and analysis in population genomics: methods and protocols*. Methods in Molecular Biology series 888. Humana Press, New York, pp. 177-196.

Denduangboripant J. & Cronk Q. C. 2001. Evolution and alignment of the hypervariable arm 1 of *Aeschynanthus* (Gesneriaceae) ITS 2 nuclear ribosomal DNA. *Molecular Phylogenetics and Evolution* 20: 163-172.

Deyoung R. W. & Honeycutt R. L. 2005. The molecular toolbox: genetic techniques in wildlife ecology and management. *Journal of Wildlife Management* 69: 1362-1384.

Deza M. M. & Deza E. 2009. *Encyclopedia of Distances*. Springer, New York.

Díaz M. L. 2004. *Comprender y proteger Sierra Nevada. Guía para divulgar sus valores ecológicos y culturales*. Centro Innovación Educativa Huerto Alegre, Granada.

Dickerson B. R., Ream R. R., Vignieri S. N. & Bentzen P. 2010. Population structure as revealed by mtDNA and microsatellites in northern fur seals, *Callorhinus ursinus*, throughout their range. *PLoS ONE* 5: e10671.

Dlugosch M. K. & Parker M. I. 2008. Founding events in species invasions: genetic variation, adaptive evolution, and the role of multiple introductions. *Molecular Ecology* 17: 431-449.

Doi H., Takahashi M. & Katano I. 2010. Genetic diversity increases regional variation in phenological dates in response to climate change. *Global Change Biology* 16: 373-379.

Douzery E. J. P., Pridgeon A. M., Kores P., Linder H. P., Kurzweil H. & Chase M. W. 1999. Molecular phylogenetics of Diseae (Orchidaceae): a contribution from nuclear ribosomal ITS sequences. *American Journal of Botany* 86: 887-899.

Downie S., Ramanath S., Katz-Downie D. & Llanas E. 1998. Molecular systematics of Apiaceae subfamily Apioideae: phylogenetic analyses of nuclear ribosomal DNA internal transcribed spacer and plastid RPO C1 intron sequences. *American Journal of Botany* 85: 563-591.

Drummond A. J., Ashton B., Buxton S., Cheung M., Cooper A., Duran C., Field M., Heled J., Kearse M., Markowitz S., Moir R., Stones-Havas S., Sturrock S., Thierer T. & Wilson A. 2011. Research software for biologists, not computer scientists. http://www.geneious.com. Accessed the 5th February 2012.

D'Souza J. S. & Johri M. M. 1999. Ca^{2+}dPKs from the protonema of the moss *Funaria hygrometrica*: effect of indole-acetic acid and cultural parameters on the activity of a 44 kDa Ca^{2+}dPK. *Plant Science* 145: 23-32.

D'Souza J. S. & Johri M. M. 2003. Purification and characterization of a Ca^{2+}-dependent/calmodulin-stimulated protein kinase from moss chloronema cells. *Journal of Bioscience* 28: 223-233.

During H. J. 1979. Life strategies of bryophytes: a preliminary review. *Lindbergia* 5: 2-18.

During H. J. 1992. Ecological classifications of bryophytes and lichens. In: Bates J. W. & Farmer A. M. (eds.), *Bryophytes and Lichens in a changing environment*. Clarendon Press, Oxford, pp. 1-31.

During H. J. & Lloret F. 1996. Permanent grid studies in bryophyte communities. I. Pattern and dynamics of individual species. *Journal of the Hattori Botanical Laboratory* 79: 1-41.

Egan S. P., Nosil P. & Funk D. J. 2008. Selection and genomic differentiation during ecological speciation: isolation the contributions of host association via a comparative genome scan of *Neochlamisus bebbianae* leaf beetles. *Evolution* 62: 316-336.

Eisen J. A. 1999. Mechanistic basis for microsatellite instability. In: Goldstein D. B. & Schlotterer C. (eds.), *Microsatellites: evolution and applications.* Oxford University Press, Oxford, pp. 34-48.

Engelmann J., Rahmann S., Wolf M., Schultz J., Fritzilas E., Kneitz S., Dandekar T. & Muller T. 2009. Modeling cross-hybridization on phylogenetic rDNA microarrays increases the detection power of closely related species. *Molecular Ecology Resources* 9: 83-93.

Excoffier L., Smouse P. E. & Quattro J. M. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131: 479-491.

Excoffier L. & Lischer H. E. L. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* 10: 564-567.

Falush D., Stephens M. & Pritchard J. K. 2003. Inference of population structure: extensions to linked loci and correlated allele frequencies. *Genetics* 164:1567-1587.

Falush D., Stephens M. & Pritchard J. K. 2007. Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular Ecology Notes* 7: 574-578.

Felsenstein J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39: 783-791.

Fernandez C. C., Shevock J. R., Glazer A. N. & Thompson J. N. 2006. Cryptic species within the cosmopolitan desiccation-tolerant moss *Grimmia laevigata*. *Proceedings of the National Academy of Sciences of the United States of America* 103: 637-642.

Fink S., Fischer M. C., Excoffier L. & Heckel G. 2010. Genomic scans support repetitive continental colonization events during the rapid radiation of voles (Rodentia: *Microtus*): the utility of AFLPs versus mitochondrial and nuclear sequence markers. *Systematic Biology* 59: 548-572.

Foll M. & Gaggiotti O. E. 2006. Identifying the environmental factors that determine the genetic structure of populations. *Genetics* 174: 875-891.

Fontaneto D. 2011. *Biogeography of microscopic organisms: is everything small everywhere?* Systematics Association Special Volume Series. Cambridge University Press, New York.

Forrest L. L., Davis E. C., Long D. G., Crandall-Stotler B. J., Clark A. & Hollingsworth M. L. 2006. Unraveling the evolutionary history of the liverworts *Marchantiophyta*: multiple taxa genomes and analyses. *The Bryologist* 109: 303-334.

Frahm J.-P. 2008. Diversity dispersal and biogeography of bryophytes mosses. *Biodiversity and Conservation* 17: 277-284.

Frahm J.-P. & Ohlemüller R. 2001. Ecology of bryophytes along altitudinal and latitudinal gradients in New Zealand. Studies in austral temperate rain forest bryophytes 15. *Tropical Bryology* 20: 117-137.

Frankham R. 1996. Relationship of genetic variation to population size in wildlife. *Conservation Biology* 10: 1500-1508.

Fritsch R. 1982. Index to plant chromosome numbers – Bryophyta. *Regnum Vegetabile* 108: 1-268.

Funk D. J., Egan S. P. & Nosil P. 2011. Isolation by adaptation in *Neochlamisus* leaf beetles: host-related selection promotes neutral genomic divergence. *Molecular Ecology* 20: 4671-4682.

Gardner K. M. & Latta R. G. 2006. Identifying loci under selection across contrasting environments in *Avena barbata* using quantitative trait locus mapping. *Molecular Ecology* 15: 1321-1333.

Gee H. 2003. Evolution: ending incongruence. *Nature* 425: 782.

Gil J. A. 1976. *Estudio briosociológico de las comunidades higrohidrófilas de Sierra Nevada*. Universidad de Granada, Ph.D. thesis.

Gil J. A. 1988. Flora y vegetación briofítica de Sierra Nevada. *Monografías de Flora y Vegetación Bética* 3: 63-72.

Gillespie J. J. 2004. Characterising regions of ambiguous alignment caused by the expansion and contraction of hairpin-stem loops in ribosomal RNA molecules. *Molecular Phylogenetics and Evolution* 33: 936-943.

Giordano S., Basile A., Spagnuolo V., Reca N. C. & Cobianchi R. 2002. Modulation of protonemal morphogenesis in *Bryum capillare* and *Pleurochaete squarrosa*: a comparison with the *Funaria hygrometrica* model system. *Plant Biosystems* 1361: 101-107.

Goldstein D. B., Zhivotovsky L. A., Nayar K., Linares A. R., Cavalli-Sforza L. L. & Feldman M. W. 1996. Statistical properties of the variation at linked microsatellite loci implications for the history of human Y-chromosome. *Molecular Biology and Evolution* 13: 1213-1218.

Goldstein D. B. & Pollock D. D. 1997. Launching microsatellites: a review of mutation processes and methods of phylogenetic inference. *Journal of Heredity* 88: 335-342.

Gómez Ortiz A. (coord.) 2002. *Mapa geomorfológico de Sierra Nevada. Morfología glaciar y periglaciar / Geomorphological Map of Sierra Nevada. Glacial and periglacial Geomorphology*. Consejería de Medio Ambiente, Parque Nacional de Sierra Nevada, Parque Natural de Sierra Nevada, Universitat de Barcelona, Granada.

Griffiths A. J. F., Miller J. H., Suzuki D. T., Lewontin R. C. & Gelbart W. M. 2000. *An Introduction to Genetic Analysis, 7th edition*. W. H. Freeman, New York.

Groth-Malonek M., Pruchner D., Grewe F. & Knoop V. 2005. Ancestors of trans-splicing mitochondrial introns support serial sister group relationships of

hornworts and mosses with vascular plants. *Molecular Biology and Evolution* 22: 117-125.

Habera L., Smith N., Donahoo R. & Lamour K. 2004. Use of a single primer to fluorescently label selective amplified fragment length polymorphism reactions. *Biotechniques* 37: 902-903.

Haldane J. B. S. 1957. The cost of natural selection. *Journal of Genetics* 55: 511-524.

Hall T. A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* 41: 95-98.

Hallingbäck T. & Hodgetts N. (compilers). 2000. *Mosses, Liverworts, and Hornworts. Status Survey and Conservation Action Plan for Bryophtyes*. IUCN/SSC Bryophyte Specialist Group. IUCN, Gland, Switzerland and Cambridge.

Hamilton M. B. 1999. Four primer pairs for the amplification of chloroplast intergenic regions with intraspecific variation. *Molecular Ecology* 8: 521-523.

Hancock J. M. 1999. Microsatellites and other simple sequences: genomic context and mutational mechanisms. In: Goldstein D. B. & Schlotterer C. (eds.), *Microsatellites: evolution and applications*. Oxford University Press, Oxford, pp. 1-9.

Hayashi J. I., Tagashira Y. & Yoshida M. C. 1985. Absence of extensive recombination between inter- and intraspecies mitochondrial DNA in mammalian cells. *Experimental Cell Research* 160: 387-395.

Hardy G. H. 1908. Mendelian proportions in a mixed population. *Science* 28: 49-50.

Hartl D. L. & Clark A. G. 1997. *Principles of population genetics*. Sinauer Associates, Inc., Sunderland, Massachusetts.

Hedenäs L. 2012a. Molecular differentiation within European *Cratoneuron filicinum*, and differences from Asiatic and American populations. *Plant Systematics and Evolution* 298: 937-945.

Hedenäs L. 2012b. Global phylogeography in *Sanionia uncinata* (Amblystegiaceae: Bryophyta). *Botanical Journal of the Linnean Society* 168: 19-42.

Hedenäs L. & Eldenäs P. 2007. Cryptic speciation, habitat differentiation, and geography in *Hamatocaulis vernicosus* (Calliergonaceae, Bryophyta). *Plant Systematics and Evolution* 268: 131-145.

Hedrick P. W., Ginevan M. E. & Ewing E. P. 1976. Genetic polymorphism in heterogeneous environments. *Annual Review of Ecology and Systematics* 7: 1-32.

Herrera C. M. & Bazaga P. 2008. Population-genomic approach reveals adaptive floral divergence in discrete populations of hawk moth-pollinated violet. *Molecular Ecology* 17: 5378-5390.

Herrmann D., Poncet B. N., Manel S., Rioux D., Gielly L., Taberlet P. & Gugerli F. 2010. Selection criteria for scoring amplified fragment length polymorphisms (AFLPs) positively affect the reliability of population genetic parameter estimates. *Genome* 53: 302-310.

Hijmans R. J., Guarino L., Cruz M. & Rojas E. 2001. Computer tools for spatial analysis of plant genetic resources data: 1. DIVA-GIS. *Plant Genetic Resources Newsletter* 127: 15-19.

Hijmans R. J., Cameron S. E., Parra J. L., Jones P. G. & Jarvis A. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* 25: 1965-1978.

Hill M. O., Bell N., Bruggeman-Nannenga M. A., Brugués M., Cano M. J., Enroth J. Flatberg K. I., Frahm J.-P., Gallego M. T., Garilleti R., Guerra J., Hedenäs L., Holyoak D. T., Hyvönen J., Ignatov M. S., Lara F., Mazimpaka V., Muñoz J. & Söderström L. 2006. An annotated checklist of the mosses of Europe and Macaronesia. *Journal of Bryology* 28: 198-267.

Hock Z., Szövényi P., Schneller J. J., Tóth Z. & Urmi E. 2008. Bryophyte diaspore bank: a genetic memory? Genetic structure and genetic diversity of surface populations and diaspore bank in the liverwort *Mannia fragrans* (Aytoniaceae). *American Journal of Botany* 95: 542-548.

Hoech W. R., Blakley K. H. & Brown W. M. 1991. Heteroplasmy suggests limited biparental inheritance of *Mytilus* mitochondrial DNA. *Science* 251: 1488-1490.

Hohenlohe P. A., Catchen J. & Cresko W. A. 2012. Population genomics analysis of model and nonmodel organisms using sequenced RAD tags. In: Pompanon F. & Bonin A. (eds.), *Data production and analysis in population genomics: methods and protocols*. Methods in Molecular Biology series 888. Humana Press, New York, pp. 235-260.

Holland B. R., Clarke A. C. & Meudt H. M. 2008. Optimizing automated AFLP scoring parameters to improve phylogenetic resolution. *Systematic Biology* 57: 347-366.

Hornschuh M., Grotha R. & Kutschera U. 2002. Epiphytic bacteria associated with the bryophyte *Funaria hygrometrica*: effects of *Methylobacterium* strains on protonema development. *Plant Biology* 4: 682-287.

Huang J. & He C. 2010. Population structure and genetic diversity of *Huperzia serrata* (Huperziaceae) based on amplified fragment length polymorphism (AFLP) markers. *Biochemical Systematics and Ecology* 38: 1137-1147.

Hubisz M. J., Falush D., Stephens M. & Pritchard J. K. 2009. Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources* 9: 1322-1332.

Hutsemékers V., Risterucci A. M., Ricca M., Boles S., Hardy O. J., Shaw A. J. & Vanderpoorten A. 2008. Identification and characterization of nuclear microsatellite loci in the aquatic moss *Platyhypnidium riparioides* (Brachytheciaceae). *Molecular Ecology Resources* 8: 1130-1132.

Hutsemékers V., Vieira C. C., Ros R. M., Huttunen S. & Vanderpoorten A. 2012. Morphology informed by phylogeny reveals unexpected patterns of species differentiation in the aquatic moss *Rhynchostegium riparioides*. *Molecular Phylogenetics and Evolution* 62: 748-755.

Huttunen S., Hedenäs L., Ignatov M. S., Devos N. & Vanderpoorten A. 2008. Origin and evolution of the northern hemisphere disjunction in the moss genus *Homalothecium* (Brachytheciaceae). *American Journal of Botany* 95: 720-730.

Ignatov M. S., Afonina O. M. & Ignatova E. A. 2006. Check-list of mosses of East Europe and North Asia. *Arctoa* 15: 1-130.

James C., Lesemann S. & Down G. 2003. Modified AFLP analysis method for species with small genomes. *Plant Molecular Biology Reporter* 21: 303-307.

Jansen R. K., Raubeson L. A., Boore J. L., dePamphilis C. W., Chumley T. W., Haberle R. C., Wyman S. K., Alverson A. J., Peery R., Herman S. J., Fourcade H. M., Kuehl J. V., McNeal J. R., Leebens-Mack J. & Cui L.

2005. Methods for obtaining and analyzing whole chloroplast genome sequences. *Methods in Enzymology* 395: 348-384.

Jansen R. K., Cai Z., Raubeson L. A., Daniell H., dePamphilis C. W., Leebens-Mack J., Müller K. F., Guisinger-Bellian M., Haberle R. C., Hansen A. K., Chumley T. W., Lee S.-B., Peery R., McNeal J. R., Kuehl J. V. & Boore J. L. 2007. Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proceedings of the National Academy of Sciences of the United States of America* 104: 19369-19374.

Jarne P. & Lagoda P. J. L. 1996. Microsatellites, from molecules to populations and back. *Trends in Ecology & Evolution* 11: 424-429.

Joost S., Bonin A., Bruford M. W., Després L., Conord C., Erhardt G. & Taberlet P. 2007. A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. *Molecular Ecology* 16: 3955-3969.

Jordan W. C., Courtney M. W. & Neigel J. E. 1996. Low levels of intraspecific genetic variation at a rapidly evolving chloroplast DNA locus in North American duckweeds (Lemnaceae). *American Journal of Botany* 83: 430-439.

Jump A. S. & Peñuelas J. 2005. Running to stand still: adaptation and the response of plants to rapid climate change. *Ecology Letters* 8: 1010-1020.

Jump A. S., Marchant R. & Peñuelas J. 2009. Environmental change and the option value of genetic diversity. *Trends in Plant Sciences* 14: 51-58.

Kamisugi Y., von Stackelberg M., Lang D., Care M., Reski R., Rensing S. A. & Cuming A. C. 2008. A sequence-anchored genetic linkage map for the moss, *Physcomitrella patens*. *The Plant Journal* 56: 855-866.

Kantety R., La Rota M., Matthews D. & Sorrells M. 2002. Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. *Plant Molecular Biology* 48: 501-510.

Kao J.-S., Wu M. & Chiang Y.-M. 1990. Cloning and characterization of chloroplast ribosomal protein-encoding genes, *rpl*16 and *rps*3, of the marine macro-algae, *Gracilaria tenuistipitata*. *Gene* 90: 221-226.

Karlin E. F., Boles S. B., Ricca M., Temsch E. M., Greilhuber J. & Shaw A. J. 2009. Three-genome mosses: complex double allopolyploid origins for triploid gametophytes in *Sphagnum*. *Molecular Ecology* 18: 1439-1454.

Karrenberg S. & Favre A. 2008. Genetic and ecological differentiation in the hybridizing campions *Silene dioica* and *S. latifolia*. *Evolution* 62: 763-773.

Kelchner S. A. 2000. The evolution of non-coding chloroplast DNA and its application in plant systematics. *Annals of the Missouri Botanical Garden* 87: 482-498.

Kelchner S. A. 2002. Group II introns as phylogenetic tools: structure, function, and evolutionary constraints. *American Journal of Botany* 89: 1651-1669.

Kelchner S. A. & Clark L. G. 1997. Molecular evolution and phylogenetic utility of the chloroplast *rpl16* intron in *Chusquea* and the Bambusoideae (Poaceae). *Molecular Phylogenetics and Evolution* 8: 385-397.

Keller A., Schleicher T., Forster F., Ruderisch B., Dandekar T., Muller T. & Wolf M. 2008. ITS2 data corroborate a monophyletic chlorophycean DO-group (Sphaeropleales). *BMC Evolutionary Biology* 8: 218.

Keller A., Wolf M. & Dandekar T. 2010. Ribosomal RNA phylogenetics: the third dimension. *Biologia* 65: 388-391.

Kenrick P. R. & Crane P. R. 1997. The origin and early evolution of plants on land. *Nature* 389: 33-39.

Kimmerer R. W. 1993. Disturbance and dominance in *Tetraphis pellucida*: a model of disturbance frequency and reproductive mode. *The Bryologist* 96: 73-79.

Kirschner M. & Gerhart J. 1998. "Evolvability". *Proceedings of the National Academy of Sciences of the United States of America* 95: 8420-8427.

Klein S. B, Wallin J. M. & Buoncristiani M. R. 2003. Addressing ambient temperature variation effects on sizing precision of AmpFlSTR® Profiler Plus™ alleles detected on the ABI Prism® 310 genetic analyzer. *Forensic Science Communications* 5 (1). http://www.fbi.gov/about-us/lab/forensic-science-communications/fsc/jan2003/klein.htm/. Accessed the 15th April 2012.

Knoop V. 2004. The mitochondrial DNA of land plants: peculiarities in phylogenetic perspective. *Current Genetics* 46: 123-139.

Knoop V. 2010. Looking for sense in the nonsense: a short review of non-coding organellar DNA elucidating the phylogeny of bryophytes. *Tropical Bryology* 31: 51-60.

Knop W. 1884. Bereitung einer concentrierten Nährstofflösung für Pflanzen. *Die Landwirthschaftlichen Versuchs-Stationen* 30: 292-294.

Koetschan C., Hackl T., Müller T., Wolf M., Förster F. & Schultz J. 2012. ITS2 Database IV: Interactive taxon sampling for internal transcribed spacer 2 based phylogenies. *Molecular Phylogenetics and Evolution* 63: 585-588.

Korpelainen, H., Jägerbrand, A. & von Cräutlein, M., 2012. Genetic structure of mosses *Pleurozium schreberi* (Willd. ex Brid.) Mitt. and *Racomitrium*

lanuginosum (Hedw.) Brid. along altitude gradients in Hokkaido, Japan. *Journal of Bryology* 34: 309-312.

Košnar J., Herbstová M., Kolář F., Koutecký P. & Kučera J. 2012. A case study of intragenomic ITS variation in bryophytes: assessment of gene flow and role of polyploidy in the origin of European taxa of the *Tortula muralis* (Musci: Pottiaceae) complex. *Taxon* 61: 709-720.

Krzesłowska M. & Woźny A. 2000. Wall thickenings - moss protonema apical cell reaction to lead. *Biologia Plantarum* 43: 93-98.

Laenen B., Désamoré A., Devos N., Shaw A. J., González-Mancebo J. M., Carine M. A. & Vanderpoorten A. 2011. Macaronesia: a source of hidden genetic diversity for post-glacial recolonization of Western Europe in the leafy liverwort *Radula lindenbergiana*. *Journal of Biogeography* 38: 631-639.

Landis F.C. & Gargas A. 2007. Using ITS2 secondary structure to create species specific oligonucleotide probes for fungi. *Mycologia* 99: 681-692.

Larsson T.-B. (compiler) 2001. Biodiversity evaluation tools for European forests. *Ecological Bulletins* 50: 1-237.

Leonardía A. A. P., Kumar P. P. & Tan B. C. 2006. Development of microsatellite markers for the tropical moss, *Acanthorrhynchium papillatum*. *Molecular Ecology Notes* 6: 396-398.

Letsch H. O., Kück P., Stocsits R. R. & Misof B. 2010. The impact of rRNA secondary structure consideration in alignment and tree reconstruction: simulated data and a case study on the phylogeny of hexapods. *Molecular Biology and Evolution* 27: 2507-2521.

Levinson G. & Gutman G. A. 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Molecular Biology and Evolution* 4: 203-221.

Li W.-H. 2006. *Molecular Evolution*. Sinauer Associates, Inc., Sunderland, Massachusetts.

Li Y.-C., Korol A. B., Fahima T. & Nevo E. 2004. Microsatellites within genes: structure, function, and evolution. *Molecular Biology and Evolution* 21: 991-1007.

Librado P. & Rozas J. 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451-1452.

Ligrone R., Duckett J. G. & Renzaglia K. S. 2000. Conducting tissues and phyletic relationships of bryophytes. *Philosophical Transactions of the Royal Society* 355: 795-813.

Liston A., Robinson W. A., Oliphant J. M., Alvarez-Buylla E. R. 1996. Length variation in the nuclear ribosomal DNA internal transcribed region of non-flowering seed plants. *Systematic Botany* 21: 109-120.

Litt M. & Luty J. M. 1989. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *The American Journal of Human Genetics* 44: 397-401.

Liu Y., Budke J. M. & Goffinet B. 2012a. Phylogenetic inference rejects sporophyte based classification of the Funariaceae (Bryophyta): rapid radiation suggests rampant homoplasy in sporophyte evolution. *Molecular Phylogenetics and Evolution* 62: 130-145.

Liu Y, Moskwa N. L. & Goffinet B. 2012b. Development of eight mitochondrial markers for Funariaceae (Musci) and their amplification success in other mosses. *American Journal of Botany* 99: 62-65.

Liu Y., Ge X.-J., Sun Q.-B. & Cao T. 2010. Development of microsatellite markers for the moss *Ptychomitrium gardneri* (Ptychomitriaceae). *American Journal of Botany* 97: 14-16.

Lloret F. 1994. Gap colonization by mosses on a forest floor: an experimental approach. *Lindbergia* 19: 122-128.

Löhne C. & Borsch T. 2005. Molecular evolution and phylogenetic utility of the petD group II intron: a case study in basal angiosperms. *Molecular Biology and Evolution* 22: 317-332.

Löhne C., Borsch T. & Wiersema J. H. 2007. Phylogenetic analysis of Nymphaeales using fast-evolving and noncoding chloroplast markers. *Botanical Journal of the Linnean Society* 154: 141-163.

López-Garrido A. C. & Sanz de Galdeano C. 2000. El levantamiento Tortoniense-Cuaternario de Sierra Nevada (Granada, España): fenómenos de tectónica gravitatoria en su borde occidental. *Geogaceta* 132: 129-132.

Luikart G., England P. R., Tallman D., Jordan S. & Taberlet P. 2003. The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews Genetics* 4: 981-994.

Luzhang R., Bei A., Backstrom N., Huaxing L., Longying W., Lixun Z. & Naifa L. 2010. Phylogeographic structure and gene flow of Himalayan snowcock (*Tetraogallus himalayensis*). *Animal Biology* 60: 449-465.

Lynch M. & Milligan B. 1994. Analysis of population-genetic structure using RAPD markers. *Molecular Ecology* 3: 91-99.

Magdy M., Ros R. M. & Werner O. 2010. The genetic structure of the *Funaria hygrometrica* Hedw. populations in Sierra Nevada (Spain) using the sequence-related amplified polymorphism (SRAP) technique. 3rd

Physcomitrella Genome Workshop, September 13th to 15th, University of Freiburg, Germany.

Maggini F., Marrocco R., Gelati T. M. & De Dominicis R. I. 1998. Length and nucleotide sequences of the internal spacers of nuclear ribosomal DNA in gymnosperms and pteridophytes. *Plant Systematics and Evolution* 213: 199-205.

Manel S., Joost S., Epperson B. K., Holderegger R., Storfer A., Rosenberg M. S., Scribner K. T., Bonin A. & Fortin M. J. 2010. Perspectives on the use of landscape genetics to detect genetic adaptive variation in the field. *Molecular Ecology* 19: 3760-3772.

Manel S., Gugerli F., Thuiller W., Alvarez N., Legendre P., Holderegger R., Gielly L., Taberlet P. & IntraBioDiv Consortium. 2012. Broad-scale adaptive genetic variation in alpine plants is driven by temperature and precipitation. *Molecular Ecology* 15: 3729-3738.

Manen J. F. & Natali A. 1995. Comparison of the evolution of ribulose-1, 5-biphosphate carboxylase (*rbc*L) and *atp*B-*rbc*L noncoding spacer sequences in a recent plant group, the tribe Rubieae (Rubiaceae). *Journal of Molecular Evolution* 41: 920-927.

Maynard Smith J. M. & Haigh J. 1974. Hitchhiking effect of a favorable gene. *Genetics Research* 23: 23-55.

Mayol M. & Rosselló J. A. 2001. Why nuclear ribosomal DNA spacers (ITS) tells different histories in *Quercus*. *Molecular Phylogenetics and Evolution* 19: 167-176.

McCauley D. E., Sundby A. K., Bailey M. F. & Welch M. E. 2007. Inheritance of chloroplast DNA is not strictly maternal in *Silene vulgaris*

(Caryophyllaceae): evidence from experimental crosses and natural populations. *American Journal of Botany* 94: 1333-1337.

McDonald D. B. & Potts W. K. 1997. DNA microsatellites as genetic markers for several scales. In: Mindell D. P. (ed.), *Avian molecular evolution and systematics*. Academic Press, San Diego, pp. 29-49.

McMahon M. M. & Sanderson M. J. 2006. Phylogenetic supermatrix analysis of GenBank sequences from 2228 papilionoid legumes. *Systematic Biology* 55: 818-836.

Meyer C. L., Vitalis R., Saumitou-Laprade P. & Castric V. 2009. Genomic pattern of adaptive divergence in *Arabidopsis halleri*, a model species for tolerance to heavy metal. *Molecular Ecology* 18: 2050-2062.

Michel F., Kazuhiko U. & Haruo O. 1989. Comparative and functional anatomy of group II catalytic introns — a review. *Gene* 82: 5-30.

Mikulášková E., Fér T. & Kučabová V. 2012. The effect of different DNA isolation protocols and AFLP fingerprinting optimizations on error rate estimates in the bryophyte *Campylopus introflexus*. *Lindbergia* 35: 7-17.

Miller D. H. & Miller H. A. 2007. *Funaria* Hedwig. In: Flora of North America Editorial Committee (ed.), *Flora of North America. Bryophytes: Mosses*, *Part 1*. *Volume 27*. Oxford University Press, New York, pp. 182-188.

Mishler B. D. 2001. Book review, the biology of bryophytes, bryophytes aren´t just small tracheophytes. *American Journal of Botany* 88: 2129-2131.

Mohamed Rizk S. 2012. *Genetic variability of the moss Ceratodon purpureus (Hedw.) Brid. in Sierra Nevada (Spain).* Universidad de Murcia, M.Sc. thesis.

Molero Mesa J., Pérez Raya F. & Valle Tendero F. 1992. *Parque Natural de Sierra Nevada. Paisaje, fauna, flora, itinerarios*. Editorial Rueda, Madrid.

Moore M. J., Soltis P. S., Bell C. D., Burleigh J. G. & Soltis D. E. 2010. Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. *Proceedings of the National Academy of Sciences of the United States of America* 107: 4623-4628.

Moritz C. 1999. Conservation units and translocations: strategies for conserving evolutionary processes. *Hereditas* 130: 217-228.

Mráz P., Gaudeul M., Rioux D., Gielly L., Choler P., Taberlet P. & IntraBioDiv Consortium 2007. Genetic structure of *Hypochaeris uniflora* (Asteraceae) suggests vicariance in the Carpathians and rapid post-glacial colonization of the Alps from an eastern Alpine refugium. *Journal of Biogeography* 34: 2100-2114.

Muirhead C. A. 2001. Consequences of population structure on genes under balancing selection. *Evolution* 55: 1532-1541.

Müller F. 2009. An updated checklist of the mosses of Chile. *Archive for Bryology* 58: 1-124.

Müller K. 2005. SeqState—primer design and sequence statistics for phylogenetic DNA data sets. *Applied Bioinformatics* 4: 65-69.

Müller T., Philippi N., Dandekar T., Schultz J. & Wolf M., 2007. Distinguishing species. *RNA* 13: 1469-1472.

Murashige T. & Skoog F. 1962. A revised medium for rapid growth and bio assays with tobacco tissue cultures. *Physiologia Plantarum* 15: 473-497.

Nakosteen P. C. & Hughes K. W. 1978. Sexual life cycle of three species of *Funariaceae* in culture. *The Bryologist* 81: 307.

Natcheva R. & Cronberg N. 2007. Maternal transmission of cytoplasmic DNA in interspecific hybrids of peat mosses, Sphagnum (Bryophyta). *Journal of Evolutionary Biology* 20: 1613-1616.

Nei M. 1987. *Molecular Evolutionary Genetics.* Columbia University Press, New York.

Nieto Feliner G. & Rosselló J. A. 2007. Better the devil you know? Guidelines for insightful utilization of nrDNA ITS in species-level evolutionary studies in plants. *Molecular Phylogenetics and Evolution* 44: 911-919.

Nishiyama T., Fujita T., Shin-I T., Seki M., Nishide H., Uchiyama I., Kamiya A., Carninci P., Hayashizaki Y., Shinozaki K., Kohara Y. & Hasebe M. 2003. Comparative genomics of *Physcomitrella patens* gametophytic transcriptome and *Arabidopsis thaliana*: implication for land plant evolution. *Proceedings of the National Academy of Sciences of the United States of America* 100: 8007-8012.

Nosil P., Egan S. P. & Funk D. J. 2008. Heterogeneous genomic differentiation between walking-stick ecotypes: “isolation by adaptation” and multiple roles for divergent selection. *Evolution* 62: 316-336.

Nosil P., Funk D. J. & Ortiz-Barrientos D. 2009. Divergent selection and heterogeneous genomic divergence. *Molecular Ecology* 18: 375-402.

Ochyra R., Bednarek-Ochyra H. & Lewis Smith R. I. 2002. New and rare moss species from subantarctic South Georgia. *Nova Hedwigia* 74: 121-147.

Ochyra R., Lewis Smith R. I. & Bednarek-Ochyra H. 2008. *The Illustrated Moss Flora of Antarctica*. Cambridge University Press, Cambridge.

Ohsawa T., Tsuda Y., Saito Y., Sawada H. & Ide Y. 2007. Altitudinal genetic diversity and differentiation of *Quercus crispula* in the Chichibu Mountains, central Japan. *International Journal of Plant Sciences* 168: 333-340.

Olmstead R. G. & Palmer J. D. 1994. Chloroplast DNA systematics: a review of methods and data analysis. *American Journal of Botany* 81: 1205-1224.

Olson M. S. & McCauley D. E. 2002. Mitochondrial DNA diversity, population structure, and gender association in the gynodioecious plant *Silene vulgaris*. *Evolution* 56: 253-262.

Olsson S., Buchbender V., Enroth J., Hedenäs L., Huttunen S. & Quandt D. 2009. Evolution of the Neckeraceae (Bryophyta): resolving the backbone phylogeny and identifying ancestral character states. *Systematics and Biodiversity* 7: 419-432.

O'Shea B. J. 2006. Checklist of the mosses of sub-Saharan Africa (version 5, 12/06). *Tropical Bryology Research Reports* 6: 1-252.

Parihar N. S. 1961. *An Introduction to Embryophyta Volume 1: Bryophyta 4th edition.* Book Depot, Allahabad.

Paris M., Boyer S., Bonin A., Collado A., David J. P. & Despres L. 2010. Genome scan in the mosquito *Aedes rusticus*: population structure and detection of positive selection after insecticide treatment. *Molecular Ecology* 19: 325-337.

Park S. D. E. 2001. *Trypanotolerance in West African cattle and the population genetic effect of selection*. University of Dublin, Ph.D. thesis.

Parks M., Cronn R. & Liston A. 2009. Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. *BioMed Central Biology* 7: 84.

Peakall R. & Smouse P. E. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* 6: 288-295.

Pedersen N. & Hedenäs L. 2003. Phylogenetic investigations of a well supported clade within the acrocarpous moss family Bryaceae: evidence from seven chloroplast DNA sequences and morphology. *Plant Systematics and Evolution* 240: 115-132.

Pedersen N., Cox C. J. & Hedenäs L. 2003. Phylogeny of the moss family Bryaceae inferred from chloroplast DNA sequences and morphology. *Systematic Botany* 28: 471-482.

Petit R. J., Csaikl U. M., Bordács S., Burg K., Coart E., Cottrell J., van Dam B., Deans J. D., Dumolin-Lapègue S., Fineschi S., Finkeldey R., Gillies A., Glaz I., Goicoechea P. G., Jensen J. S., König A. O., Lowe A. J., Madsen S. F., Mátyás G., Munro R. C., Olalde M., Pemonge M.-H., Popescu F., Slade D., Tabbener H., Taurchini D., de Vries S. G. M., Ziegenhagen B. & Kremer A. 2002. Chloroplast DNA variation in European white oaks: phylogeography and patterns of diversity based on data from over 2600 populations. *Forest Ecology and Management* 156: 5-26.

Pigliucci M. 2001. *Phenotypic plasticity: beyond nature and nurture*. Johns Hopkins University Press, Baltimore.

Pigliucci M. & Murren C. J. 2003. Genetic assimilation and a possible evolutionary paradox: can macroevolution sometimes be so fast as to pass us by? *Evolution* 57: 1455-1464.

Pisa S., Ros R. M. & Werner O. 2010. Genetic diversity of *Bryum argenteum* Hedw. along an altitudinal gradient in Sierra Nevada, Spain. 3rd *Physcomitrella* Genome Workshop, September 13th to 15th, University of Freiburg, Germany.

Pompanon F., Bonin A., Bellemain E. & Taberlet P. 2005. Genotyping errors: causes, consequences and solutions. *Nature Reviews Genetics* 6: 847-859.

Pompanon F. & Bonin A. 2012. *Data production and analysis in population genomics: methods and protocols*. Methods in Molecular Biology series 888, Humana Press, New York.

Poncet B. N., Herrmann D., Gugerli F., Taberlet P., Holderegger R., Gielly L., Rioux D., Thuiller W., Aubert S. & Manel S. 2010. Tracking genes of ecological relevance using a genome scan in two independent regional population samples of *Arabis alpina*. *Molecular Ecology* 19: 2896-2907.

Porter C. H. & Collins F. H. 1991. Species-diagnostic differences in the ribosomal DNA internal transcribed spacer from the sibling species *Anopheles freeborni* and *Anopheles hermsi* (Diptera: Culicidae). *The American Journal of Tropical Medicine and Hygiene* 45: 271-279.

Porter S. S. & Rice K. J. 2013. Trade-offs, spatial heterogeneity, and the maintenance of microbial diversity. *Evolution* 67: 599-608.

Posada D. 2008. jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution* 25: 1253-1256.

Powell W., Morgante M., McDevitt R., Vendramin G. G. & Rafalski J. A. 1995. Polymorphic simple sequence repeat regions in chloroplast genomes: applications to the population genetics of pines. *Proceedings of the National Academy of Sciences of the United States of America* 92: 7759-7763.

Price T. D., Qvarnstrom A. & Irwin D. E. 2003. The role of phenotypic plasticity in driving genetic evolution. *Proceedings of the Royal Society of London, Series B,* 270: 1433-1440.

Prieto Fernández P. 1975. *Flora de la Tundra de Sierra Nevada*. Universidad de Granada, Granada.

Pritchard J. K., Stephens M., Rosenberg N. A. & Donnelly P. 2000a. Association mapping in structured populations. *American Journal of Human Genetics* 67: 170-181.

Pritchard J. K., Stephens M. & Donnelly P. 2000b. Inference of population structure using multilocus genotype data. *Genetics* 155: 945-959.

Proctor M. C. F. 2000. Physiological ecology. In: Shaw A. J. & Goffinet B. (eds.), *Bryophyte Biology*. Cambridge University Press, Cambridge, pp. 225-448.

Provan J. & Wilson P. J. 2007. Development of microsatellites for the peat moss *Sphagnum capillifolium* using ISSR cloning. *Molecular Ecology Notes* 7: 254-256.

Qiu Y.-L., Cho Y., Cox J. C. & Palmer J. D. 1998: The gain of three mitochondrial introns identifies liverworts as the earliest land plants. *Nature* 394: 671-674.

Qiu Y.-L., Li L., Wang B., Chen Z., Knoop V., Groth-Malonek M., Dombrovska O., Lee J., Kent L., Rest J., Estabrook G. F., Hendry T. A., Taylor D. W., Testa C. M., Ambros M., Crandall-Stotler B., Duff R. J., Stech M., Frey W., Quandt D. & Davis C. C. 2006. The deepest divergences in land plants inferred from phylogenomic evidence. *Proceedings of the National Academy of Sciences of the United States of America* 103: 15511-15516.

Qiu Y.-L., Li L., Wang B., Xue J.-Y., Hendry T. A., Li R.-Q., Brown J. W., Liu Y., Hudson G. T. & Chen Z.-D. 2010. Angiosperm phylogeny inferred from sequences of four mitochondrial genes. *Journal of Systematics and Evolution* 48: 391-425.

Quandt D. & Stech M. 2004. Molecular evolution of the *trn*T_{UGU}-*trn*F_{GAA} region in bryophytes. *Plant Biology* 6: 545-554.

Queller D. C., Strassman J. E. & Hughes C. R. 1993. Microsatellites and kinship. *Trends in Ecology and Evolution* 8: 285-288.

Rambaut A & Drummond A. J. 2007. Tracer v1.4. http://beast.bio.ed.ac.uk/Tracer. Accessed the 11th June 2012.

Rams S., Ros R. M., Cano M. J. & Guerra J. 2001. Checklist de los briófitos de Sierra Nevada (Andalucia, España). *Boletin de la Sociedad Española de Briología* 18/19: 137-164.

Rams Sánchez S. 2007. *Estudios briológicos sobre flora, vegetación, taxonomía y conservación en Sierra Nevada (Andalucía, S de España)*. Universidad de Murcia, Ph.D. thesis.

Reed D. H. & Frankham R. 2003. Correlation between fitness and genetic diversity. *Conservation Biology* 17: 230-237.

Rensing S. A., Rombauts S., van de Peer Y. *&* Reski R. 2002. Moss transcriptome and beyond*. Trends in Plant Science* 7: 535-538*.*

Rensing S. A., Lang D., Zimmer A. D., Terry, A., Salamov A., Shapiro H., Nishiyama T., Perroud P.-F., Lindquist E. A., Kamisugi Y., Tanahashi T., Sakakibara K., Fujita T., Oishi K., Shin-I T., Kuroki Y., Toyoda A., Suzuki Y., Hashimoto S.-i., Yamaguchi K., Sugano S., Kohara Y., Fujiyama A., Anterola A., Aoki S., Ashton N., Barbazuk W. B., Barker E., Bennetzen J. L., Blankenship R., Cho S. H., Dutcher S. K., Estelle M., Fawcett J. A., Gundlach H., Hanada K., Heyl A., Hicks K. A., Hughes J., Lohr M., Mayer K., Melkozernov A., Murata T., Nelson D. R., Pils B., Prigge M., Reiss B., Renner T., Rombauts S., Rushton P. J., Sanderfoot A., Schween G., Shiu S.-H., Stueber K., Theodoulou F. L., Tu H., Van de Peer Y., Verrier P. J., Waters E., Wood A., Yang L., Cove D., Cuming A. C., Hasebe M., Lucas S., Mishler B. D., Reski R., Grigoriev I. V., Quatrano R. S. & Boore J. L.

2008. The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. *Science* 319: 64-69.

Richardson B., Rehfeldt G. & Kim M. 2009. Congruent climate related genecological responses from molecular markers and quantitative traits for western white pine (*Pinus monticola*). *International Journal of Plant Sciences* 170: 1120-1131.

Rivas-Martínez S. 1991. Bioclimatic belts of West Europe (relations between bioclimate and plant ecosystems). *Proceedings of the European School of Climatology and Natural Hazards course* (Arles 1990): 225-246.

Robinson B. W. & Dukas R. 1999. The influence of phenotypic modifications on evolution: the Baldwin effect and modern perspectives. *Oikos* 85: 582-589.

Ronquist F., Teslenko M., van der Mark P., Ayres D., Darling A., Höhna S., Larget B., Liu L., Suchard M. A. & Huelsenbeck J. P. 2011. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* 61: 539-542.

Ros R. M., Mazimpaka V., Abou-Salama U., Aleffi M., Blockeel T. L., Brugués M., Cros R. M., Dia M. G., Dirkse G. M., Draper I., El Saadawi W., Erdağ A., Ganeva A., Gabriel R., González-Mancebo J. M., Granger C., Herrnstadt I., Hugonnot V., Khalil K., Kürschner H., Losada-Lima A., Luís L,, Mifsud S., Privitera M., Puglisi M., Refai M. S., Sabovljević M., Sérgio C., Shabbara H., Sim-Sim M., Sottiaux A., Tacchi R., Vanderpoorten A. & Werner O. 2013. Mosses of the Mediterranean, an annotated checklist. *Cryptogamie*, *Bryologie* 34: 99-283.

Rossi M., Bitocchi E., Bellucci E., Nanni L., Rau D., Attene G. & Papa R. 2009. Linkage disequilibrium and population structure in wild and domesticated populations of *Phaseolus vulgaris* L. *Evolutionary Applications* 2: 504-522.

Royal Botanic Gardens of Kiew. 2012. DNA barcoding protocols: RNA polymerase Rpb1 (*rpoC*1) gene primers. http://www.kew.org/barcoding/protocols.html. Accessed the 25th May 2011.

Royal Botanic Gardens Melbourne 2012. *Funaria* Hedw. www.rbg.vic.gov.au/. Accessed the 10th June 2012.

Rüdinger M., Szövényi P., Rensing S. A. & Knoop V. 2011. Assigning DYW-type PPR proteins to RNA editing sites in the funariid mosses *Physcomitrella patens* and *Funaria hygrometrica*. *Plant Journal* 67: 370-380.

Russell P. 2001. *iGenetics*. Benjamin Cummings, New York.

Rütten D. & Santarius K. A. 1992. Relationship between frost tolerance and sugar concentration of various bryophytes in summer and winter. *Oecologia* 91: 260-265.

Rütten D. & Santarius K. A. 1993. Seasonal variation in frost tolerance and sugar content of two *Plagiomnium* species. *The Bryologist* 96: 564-568.

Sabovljević M., Bijelović A. & Dragičević I. 2003. In vitro culture of mosses: *Aloina aloides* (K. F. Schultz) Kindb., *Brachythecium velutinum* (Hedw.) B. S. & G., *Ceratodon purpureus* (Hedw.) Brid., *Eurhynchium praelongum* (Hedw.) B. S. & G. and *Grimmia pulvinata* (Hedw.) Sm. *Turkish Journal of Botany* 27: 441-446.

Sakai M., Kanazawa A., Fujii A., Thseng F. S., Abe J. & Shimamoto Y. 2003. Phylogenetic relationships of the chloroplast genomes in the genus *Glycine* inferred from four intergenic spacer sequences. *Plant Systematics and Evolution* 239: 29-54.

Sakakibara H. 2006. Cytokinins: activity biosynthesis and translocation. *Annual Review of Plant Biology* 57: 431-49.

Sambrook J., Fritsch E. F. & Maniatis T. 1989. *Molecular cloning: a laboratory manual. 2nd edition.* Cold Spring Harbor Laboratory Press, New York.

Sang T., Crawford D. J. & Stuessy T. F. 1997. Chloroplast DNA phylogeny, reticulate evolution, and biogeography of *Paeonia* (Paeoniaceae). *American Journal of Botany* 84: 1120-1136.

Savolainen V., Anstett M. C., Lexer C., Hutton I., Clarkson J. J., Norup M. V., Powell M. P., Springate D., Salamin N. & Baker W. J. 2006. Sympatric speciation in palms on an oceanic island. *Nature* 441: 210-213.

Scarcelli N., Barnaud A., Eiserhardt W., Treier U. A., Seveno M., d'Anfray A., Vigouroux Y. & Pintaud J.-C. 2011. A set of 100 chloroplast DNA primer pairs to study population genetics and phylogeny in monocotyledons. *PLoS ONE* 6: e19954.

Schaefer D. G. 2001. Gene targeting in *Physcomitrella patens*. *Current Opinion in Plant Biology* 4: 143-150.

Schaefer D. G., Zrÿd J.-P., Knight C. D. & Cove D. J. 1991. Stable transformation of the moss *Physcomitrella patens*. *Molecular and General Genetics* 226: 418-424.

Schaefer D. G. & Zrÿd J. P. 1997. Efficient gene targeting in the moss *Physcomitrella patens*. *Plant Journal* 11: 1195-1206.

Schaefer D. G. & Zrÿd J. P. 2001. The moss *Physcomitrella patens* now and then. *Plant Physiology* 127: 1430-1438.

Scherrer D. & Körner C. 2011. Topographically controlled thermal-habitat differentiation buffers alpine plant diversity against climate warming. *Journal of Biogeography* 38: 406-416.

Schierup M. H. 1998. The number of self-incompatibility alleles in a finite, subdivided population. *Genetics* 149: 1153-1162.

Schlichting C. D. 2004. The role of phenotypic plasticity in diversification. In: DeWitt T. J. & Scheiner S. M. (eds.), *Phenotypic plasticity: functional and conceptual approaches.* Oxford University Press, Oxford, pp. 191-200.

Schlotterer C. 1998. Microsatellites. In: Hoelzel A. R. (ed.), *Molecular genetic analysis of populations: a practical approach*. IRL Press, Oxford, pp. 237-261.

Schlötterer C. 2003. Hitchhiking mapping-functional genomics from the population genetics perspective. *Trends in Genetics* 19: 32-38.

Schlüter P. M. & Harris S. A. 2006. Analysis of multilocus fingerprinting data sets containing missing data. *Molecular Ecology Notes* 6: 569-572.

Schofield W. B. 1985. *Introduction to Bryology*. Macmillan Publishing Company, New York.

Schultz J. & Wolf M. 2009. ITS2 sequence–structure analysis in phylogenetics: a how-to manual for molecular systematics. *Molecular Phylogenetics and Evolution* 52: 520-523.

Seppelt R. D. 2004. *The moss flora of Macquarie Island*. Australian Antarctic Division, Kingston.

Sharp A. J., Crum H. A. & Eckel P. M. 1994. *The Moss Flora of Mexico*. New York Botanical Garden, New York.

Shaw A. J. 1990. Metal tolerances and co-tolerances in the moss *Funaria hygrometrica*. *Canadian Journal of Botany* 68: 2275-2282.

Shaw A. J. 1991. The genetic structure of sporophytic and gametophytic populations of the moss *Funaria hygrometrica* Hedw. *Evolution* 455: 1260-1274.

Shaw A. J. & Bartow S. M. 1992. Genetic structure and phenotypic plasticity in proximate populations of the moss *Funaria hygrometrica*. *Systematic Botany* 172: 257-271.

Shaw A. J., Werner O. & Ros R. M. 2003. Intercontinental Mediterranean disjunct mosses: morphological and molecular patterns. *American Journal of Botany* 90: 540-550.

Shaw A. J. & Renzaglia K. S. 2004. Phylogeny and diversification of bryophytes. *American Journal of Botany* 91: 1557-1581.

Simmons M. P. & Ochoterena H. 2000. Gaps as characters in sequence-based phylogenetic analyses. *Systematic Biology* 49: 369-381.

Sitte P., Ziegler H., Ehrendorfer F. & Bresinsky A. 1994. *Strasburger Tratado de Botánica, 8ª edición castellana*. Omega, Barcelona.

Smith A. J. E. 2004. *The Moss Flora of Britan and Ireland*, *Second edition*. Cambridge University Press, Cambridge.

Soltis D. E., Mavrodiev E. V., Doyle J. J., Rauscher J. & Soltis P. S. 2008. ITS and ETS sequence data and phylogeny reconstruction in allopolyploids and hybrids. *Systematic Botany* 33: 7-20.

Song N., Nwafili S. A. & Gao T. X. 2011. Genetic diversity and population structure of *Chrysichthys nigrodigitatus* from Niger Delta based on AFLP analysis. *Biochemical Systematics and Ecology* 39: 320-327.

Sonibare M. A., Asiedu R. & Albach D. C. 2010. Genetic diversity of *Dioscorea dumetorum* (Kunth) Pax using Amplified Fragment Length Polymorphisms (AFLP) and cpDNA. *Biochemical Systematics and Ecology* 38: 320-334.

Sousa P., Froufe E., Harris D. J., Alves P. C. & van der Meijden A. 2011. Genetic diversity of Maghrebian *Hottentotta* (Scorpiones: Buthidae) scorpions based on CO1: new insights on the genus phylogeny and distribution. *African Invertebrates* 52: 135-143.

Stanton M. L. & Galen C. 1997. Life on the edge: adaptation versus environmentally mediated gene flow in the snow buttercup, *Ranunculus adoneus*. *American Naturalist* 150: 143-178.

Stech M., Quandt D. & Frey W. 2003. Molecular circumscription of the hornworts Anthocerotophyta based on the chloroplast DNA *trnL-trnF* region. *Journal of Plant Research* 116: 389-398.

Stech M. & Quandt D. 2010. 20,000 species and five key markers: the status of molecular bryophyte phylogenetics. *Phytotaxa* 9: 196-228.

Stech M., Werner O., González-Mancebo J. M., Patiño J., Sim-Sim M., Fontinha S., Hildebrandt I. & Ros R. M. 2011. Phylogenetic inference in *Leucodon* Schwägr. subg. *Leucodon* (Leucodontaceae, Bryophyta) in the North Atlantic region. *Taxon* 60: 79-88.

Stenøien H. K. 1999. Are enzyme loci selectively neutral in haploid populations of nonvascular plants? *Evolution* 53: 1050-1059.

Stenøien H. K. & Såstad S. M. 2001. Genetic variation in bryophytes: does mating system really matter? *Journal of Bryology* 23: 313-318.

Stephan W., Xing L., Kirby D. A. & Braverman J. M. 1998. A test of the background selection hypothesis based on nucleotide data from

Drosophila ananassae. *Proceedings of the National Academy of Sciences of the United States of America* 95: 5649-5654.

Stinson K. A. 2004. Natural selection favors rapid reproductive phenology in *Potentilla pulcherrima* (Rosaceae) at opposite ends of a subalpine snowmelt gradient. *American Journal of Botany* 91: 531-539.

Storz J. F. 2005. Using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular Ecology* 14: 671-688.

Sundberg S. & Rydin H. 2002. Habitat requirements for establishment of *Sphagnum* from spores. *Journal of Ecology* 90: 268-278.

Svab Z. & Maliga P. 2007. Exceptional transmission of plastids and mitochondria from the transplastomic pollen parent and its impact on transgene containment. *Proceedings of the National Academy of Sciences of the United States of America* 104: 7003-7008.

Swofford D. L. 2002. PAUP* Phylogenetic analysis using parsimony (*and other methods). Sinauer Associates, Inc., Sunderland, Massachusetts.

Szövényi P., Rensing S. A., Lang D., Wray G. A. & Shaw A. J. 2010. Generation-biased gene expression in a bryophyte model system. *Molecular Biology and Evolution* 28: 803-812

Taberlet P., Gielly L., Patou G. & Bouvet J. 1991. Universal primers for amplification of three noncoding regions of chloroplast DNA. *Plant Molecular Biology* 17: 1105-1109.

Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123: 585-595.

Tamura K., Peterson D., Peterson N., Stecher G., Nei M. & Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum

likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* 28: 2731-2739.

Tanahashi T., Sumikawa N., Kato M. & Hasebe M. 2005. Diversification of gene function: homologs of the floral regulator FLO/LFY control the first zygotic cell division in the moss *Physcomitrella patens*. *Development* 132: 1727-1736.

Tautz D., Trick M. & Dover G. A. 1986. Cryptic simplicity in DNA is a major source of genetic variation. *Nature* 322: 652-656.

Taylor P. J., Eppley S. M. & Jesson L. K. 2007. Sporophytic inbreeding depression in mosses occurs in a species with separate sexes but not in a species with combined sexes. *American Journal of Botany* 94: 1853-1859.

Teacher A. G. F. & Griffiths D. J. 2011. HapStar: automated haplotype network layout and visualization. *Molecular Ecology Resources* 11: 151-153.

Terao H. 1918. Maternal inheritance in the soy bean. *The American Naturalist* 52: 51-56.

Thompson J. D., Higgins D. G. & Gibson T. J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* 22: 4673-4680.

Tietz D. 1998. *Nucleic Acid Electrophoresis Laboratory Manual*. Springer Verlag, Berlin, Heidelberg, New York.

Tollenaere C., Duplantier J. M., Rahalison L., Ranjalahy M. & Brouat C. 2011. AFLP genome scan in the black rat (*Rattus rattus*) from Madagascar: detecting genetic markers undergoing plague-mediated selection. *Molecular Ecology* 20: 1026-1038.

Tropicos 2012. Missouri Botanical Garden. http://www.tropicos.org. Accessed the 25th June 2012.

Turesson G. 1925. The plant species in relation to habitat and climate. *Hereditas* 6: 147-236.

van der Velde M., van der Strate H. J., van de Zande L. & Bijlsma R. 2000. Isolation and characterization of microsatellites in the moss species *Polytrichum formosum*. *Molecular Ecology* 9: 1678-1680.

van Reenen G. B. A. & Gradstein S. R. 1983a. Studies on Colombian Crytogams XX. A transect analysis of the bryophyte vegetation along an altitudinal gradient on the Sierra Nevada de Santa Marta, Colombia. *Acta Botanica Neerlandica* 32: 163-175.

van Reenen G. B. A. & Gradstein S. R. 1983b. An investigation of bryophyte distribution and ecology along an altitudinal gradient in the Andes of Colombia. *Journal of the Hattori Botanical Laboratory* 55: 79-84.

Vanderpoorten A. & Goffinet B. 2009. *Introduction to bryophytes*. Cambridge University Press, Cambridge.

Vekemans X., Beauwens T., Lemaire M. & Roldan-Ruiz I. 2002. Data from amplified fragment length polymorphism (AFLP) markers show indication of size homoplasy and of a relationship between degree of homoplasy and fragment size. *Molecular Ecology* 11: 139-151.

von Schwartzenberg K. 2006. Moss biology and phytohormones - cytokinins in *Physcomitrella*. *Plant Biology* 8: 382-388.

Vos P., Hogers R., Bleeker M., Reijans M., Lee T.v.d., Horne, M., Friters A., Pot J., Paleman J., Kuiper M. & Zabeau M. 1995. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research* 23: 4407-4414.

Wang R. H., Farrona S., Vincent C., Joecker A., Schoof H., Turck F., Alonso-Blanco C., Coupland G. & Albani M. C. 2009. PEP1 regulates perennial flowering in *Arabis alpina*. *Nature* 459: 423-427.

Wen C. S. & Hsiao J. Y. 2001. Altitudinal genetic differentiation and diversity of Taiwan lily (*Lilium longiflorum* var. *formosanum*; Liliaceae) using RAPD markers and morphological characters. *International Journal of Plant Sciences* 162: 287-295.

Werner O., Rams S. & Ros R. M. 2005. Genetic diversity of *Pohlia bolanderi* (Mniaceae), a rare and threatened moss in Sierra Nevada (Spain), estimated by ISSR molecular markers. *Nova Hedwigia* 81: 413-420.

Werner O., Ros R. M. & Goffinet B. 2007. A reconsideration of the systematic position of *Goniomitrium* (Funariaceae) based on chloroplast sequence markers. *The Bryologist* 110: 108-114.

Werner O., Patiño J., González–Mancebo J. M., de Almeida Gabriel R. M. & Ros R. M. 2009. The taxonomic status and the geographical relationships of the Macaronesian endemic moss *Fissidens luisieri* (Fissidentaceae) based on DNA sequence data. *The Bryologist* 112: 315-324.

West-Eberhard M. J. 2003. *Developmental Plasticity and Evolution*. Oxford University Press, New York.

Wheeler W. C. & Honeycutt R. L. 1988. Paired sequence difference in ribosomal RNAs: evolutionary and phylogenetic implications. *Molecular Biology and Evolution* 5: 90-96.

White T., Bruns T., Lee S. & Taylor J. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis M., Gelfand D., Shinsky J. & White T. (eds.), *PCR Protocols: a guide to methods and applications.* Academic Press, Inc., Michigan, pp. 315-322.

Whiteley A. R., Spruell P. & Allendorf F. W. 2006. Can common species provide valuable information for conservation? *Molecular Ecology* 15: 2767-2786.

Wilding C. S., Butlin R. K. & Grahame J. 2001. Differential gene exchange between parapatric morphs of *Littorina saxatilis* detected using AFLP markers. *Journal of Evolutionary Biology* 14: 611-619.

Willing E. M., Bentzen P., van Oosterhout C., Hoffmann M., Cable J., Breden F., Weigel D. & Dreyer C. 2010. Genome-wide single nucleotide polymorphisms reveal population history and adaptive divergence in wild guppies. *Molecular Ecology* 19: 968-984.

Wolf J. H. D. 1993. Diversity patterns and biomass of epiphytic bryophytes and lichens along an altitudinal gradient in the northern Andes. *Annals of the Missouri Botanical Garden* 80: 928-960.

Won H. & Renner S. S. 2005. The chloroplast *trn*T-*trn*F region in the seed plant lineage gnetales. *Journal of Molecular Evolution* 61: 425-436.

WorldClim - Global Climate Data. 2012. Free climate data for ecological modeling and GIS. http://www.worldclim.org/download. Accessed the 10th October 2012.

Wright S. 1931. Evolution in Mendelian populations. *Genetics* 16: 97-159.

Wright S. 1978. *Evolution and the genetics of populations. Volume 4. Variability within and among natural populations.* University of Chicago press, Chicago.

Yang T. W., Yang Y. A. & Xiong Z. 2000. Paternal inheritance of chloroplast DNA in interspecific hybrids in the genus *Larrea* (Zygophyllaceae). *American Journal of Botany* 87: 1452-1458.

Yeh F. C. & Boyle T. J. B. 1997. Population genetic analysis of co-dominant and dominant markers and quantitative traits. *Belgian Journal of Botany* 129: 157.

Zane L., Bargelloni L. & Patarnello T. 2002. Strategies for microsatellite isolation: a review. *Molecular Ecology* 11: 1-16.

Zhang Z., Schwartz S., Wagner L. & Miller W. 2000. A greedy algorithm for aligning DNA sequences. *Journal of Computational Biology* 7: 203-214.

Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research* 31: 3406-3415.