


Short Communication

Transcriptome assembly and polymorphism detection in *Silene ciliata* (Caryophyllaceae)

Sandra Sacristán-Bajo¹, Alfredo García-Fernández¹, Jose M. Iriondo¹ and Carlos Lara-Romero^{1,2*} 

¹Biodiversity and Conservation Area, Superior School of Experimental Science and Technology (ESCET), Rey Juan Carlos University, Madrid, Spain and ²Global Change Research Group, Mediterranean Institute of Advanced Studies (CSIC–IUB), Esporles, Mallorca, Spain

Received 18 January 2019; Accepted 5 April 2019

Abstract

Silene ciliata (Caryophyllaceae) is a key species to test evolutionary hypotheses in a global warming context. The recent advances in Next Generation Sequencing technologies can help in providing clues about climate-mediated local adaptation. In the present study, we analysed the full transcriptome of six individuals of *S. ciliata* from Central Spain, by aligning it with the transcriptome of *S. latifolia*. We aimed (a) to identify Single Nucleotide Polymorphisms (SNPs) in the transcriptome of the species, (b) to describe the biological function of the polymorphic genes expressed and (c) to identify loci that may be involved in local adaptation processes at optimal and marginal populations of the species. We identified a total of 147,118 SNPs distributed throughout 12,688 sequences. The number of polymorphic sequences annotated was 8023. One hundred thirty sequences containing polymorphisms strongly associated with optimal and marginal conditions were selected. Gene ontology searches were successful for 118, and many of these were related to responses to stress ($n = 19$) and abiotic stimulus ($n = 16$). Genomic data generated provide a starting point for further research on the identification of candidate genes related to local adaptation and other processes in the species.

Keywords: adaptation, alpine plants, functional annotation, SNPs

Introduction

Silene L. (Caryophyllaceae) is a key plant genus for studying crucial questions interrelating ecology and evolution (Bernasconi *et al.*, 2009). *Silene ciliata* Pourret is a Mediterranean alpine plant that occurs in alpine pastures (Escudero *et al.*, 2005) protected by the European Habitats Directive (Council Directive 92/43/EEC, 1992). The species is threatened by global warming (Giménez-Benavides *et al.*, 2007, 2018) and has been included in catalogues of

threatened species from different countries and regions (Dray, 1985; Fernández *et al.*, 2007; Sanz *et al.*, 2010; Légifrance, 2019). The seedling stage of this species experiences great mortality and it is, thus, subjected to strong selective pressure. This pressure may be qualitatively different between environmental conditions that are most commonly found at the species populations (optimal conditions) and those found only at the extreme of the species ecological range (marginal conditions). Previous studies have shown local adaptation patterns in optimal and marginal populations (Giménez-Benavides *et al.*, 2007, 2018; García-Fernández *et al.*, 2015). Thus, *S. ciliata* is a key exponent to evaluate Mediterranean alpine species responses to

*Corresponding author. E-mail: carlos.lara.romero@gmail.com

oncoming global warming and corroborate evolutionary hypotheses (e.g. García-Fernández *et al.*, 2015; Kyrkou *et al.*, 2015; Lara-Romero *et al.*, 2016). Accordingly, there is great interest in developing genomic resources for the species, which would improve the understanding of the genetics of adaptation in Mediterranean alpine environments.

In this context, we carried out a transcriptomic study of seedlings of *S. ciliata*, with the following objectives: (1) to identify Single Nucleotide Polymorphisms (SNPs) in the transcriptome of the species, (2) to describe the biological function of the polymorphic genes expressed, (3) to examine diversity patterns of potential adaptive value in optimal and marginal populations of the species and identify loci that may be involved in local adaptation processes.

Experimental

We used RNeasy Plant Mini-Kit (QIAGEN) to extract and isolate RNA from six seedlings grown under controlled conditions, one for each of the six studied populations located in Central Spain. Three populations were located at the high edge and the other three at the low edge of the species elevational range (Table 1). The high and low-elevation edges represent optimum and marginal (warmer and drier) environmental conditions for the species, respectively. The quality of RNA was evaluated with a Qubit (Invitrogen, Carlsbad, CA, USA). One sequencing run was carried out in an Illumina platform through 100 bp paired-end reads. Trimming was carried out with software Trimmomatic (Bolger *et al.*, 2014). Then, *S. ciliata* transcriptome was aligned with the genome of *Silene latifolia* (GenBank reference: GCA_900095335.1) using BWA software (Li and Durbin, 2010).

SNPs were identified using Reads2snp (Gayral *et al.*, 2013) and filtered with VCFtools 4.1 (Danecek *et al.*, 2011). Only, biallelic SNPs with no missing data and at least seven reads per genotype were retained to prevent the inclusion of false positive SNPs (Swarts *et al.*, 2014; Marano *et al.*, 2017). Paralogous and singleton SNPs were further deleted. VCFtools was also used to estimate the genetic variation in the whole genome. Blastx software (Altschul *et al.*, 1990) and the database of SWISS-PROT (Bairoch and Apweiler, 2000) were used to annotate the biological function (i.e. gene ontology terms) of the sequences carrying SNPs. We applied two different measures to detect candidate SNPs with unusually high-allele frequency differentiation between elevations. We first calculated allele frequency differences (AFDs) between low and high elevations at the individual allele level (Turner *et al.*, 2010; Stölting *et al.*, 2015). SNPs were considered unusually divergent if AFDs were ≥ 3 SDs higher than the genome-wide average. Second, following Muller *et al.*, 2011, we computed the dispersion of each allele (m),

Table 1. Geographic and climatic characteristics of the six studied populations of *Silene ciliata* in Central Spain and genetic assessment of the sampled seedlings

Locality	Elevation (m)	Env. class	Min T (°C)	Snowpack	Lat.	Long.	Herbarium sheet	Mapping (%)	Total sequences ($\times 10^6$)	H_O	F_i	Tajima's π^a
Pico de Peñalara	2400	Optimal	-7.8	0.67	40°51'2.1"N	3°57'24.1"W	MA880768	45	27.2	0.249	0.025	0.364×10^{-3}
Altos del Morezón	2380	Optimal	-7.7	0.62	40°14'57.5"N	5°16'8.3"W	MA880763	43	34.6	0.286	0.053	0.358×10^{-3}
Canchal Negro	2360	Optimal	-7.2	0.59	40°20'19.9"N	5°41'22.3"W	MA880766	44.6	27.5	0.120	0.120	0.326×10^{-3}
Las Cimeras	2000	Marginal	-6.7	0.27	40°21'7.1"N	5°40'59.7"W	MA880765	37.9	26.2	0.270	0.105	0.350×10^{-3}
Morrena Peñalara	1980	Marginal	-5.7	0.33	40°50'11.8"N	3°57'0.9"W	MA880769	40	34.6	0.309	-0.026	0.383×10^{-3}
Los Campanarios	2000	Marginal	-6.0	0.34	40°15'42.6"N	5°12'55.7"W	MA880764	38.29	34.5	0.288	0.045	0.351×10^{-3}

For each plant, the population of origin, environmental classification, minimum annual temperature in thaw months (February, March and April), geographical coordinates, percentage of mapping with the reference genome, total number of sequences retained after trimming and estimators of genetic diversity are provided. Trimming configuration after the optimization process was: leading: 5, trailing: 5, sliding window: 4:15, minlen: 50. See trimmomatic manual (<http://www.usadellab.org/cms/?page=trimmomatic>) for further details on the selection of trimming steps and their associated parameters.

H_O , observed heterozygosity; F_i , coefficient of inbreeding.

^aNucleotide diversity per site $\times 10^{-3}$. Climatic variables were obtained from the ACPI (Atlas Climatic de la Península Ibérica, <http://opengis.uab.es/wms/iberia/index.htm>). Monthly snowpack was calculated following the methodology proposed by López-Moreno and co-workers (2007) and it ranged from 0 to 1.

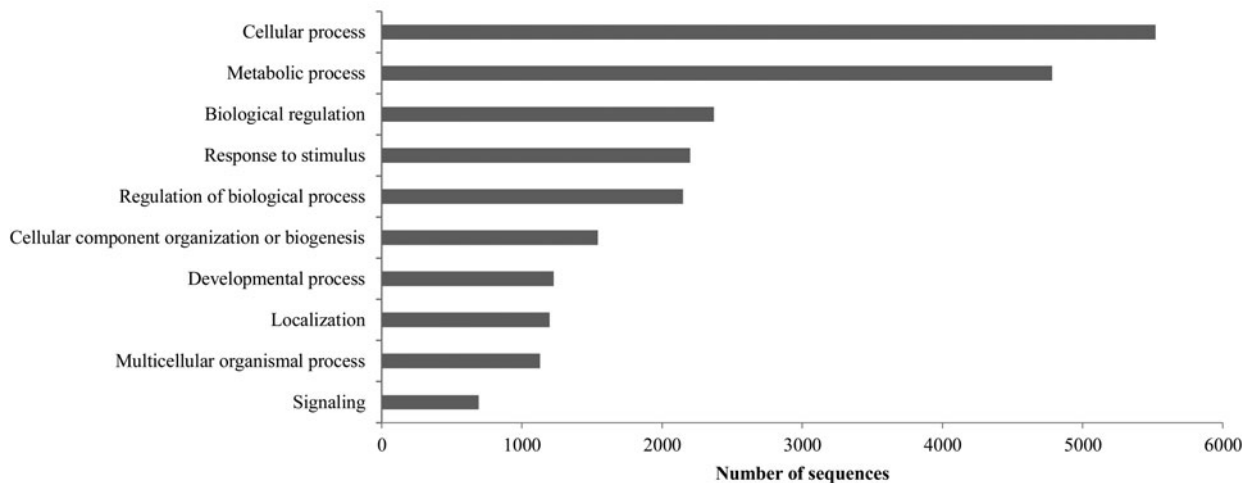


Fig 1. Most common biological functions found in polymorphic genes. The graph was created using the web Gene Ontology Annotation Plotting WEGO 2.0 (Ye *et al.*, 2018).

which is the average elevational distance of an allele copy to the average elevation of all copies of that allele (β). Then, 1000 permutations of allele copies among all studied geographical locations were performed, to subsequently estimate m and β in each of the permutations. SNPs were considered unusually clustered if m were ≥ 2 SDs higher than the permutation average. We performed a GO term enrichment analyses using Fisher's exact tests to assess whether sequences containing selected SNPs were enriched in a biological function.

Discussion

Eighty percent of RNA sequences were conserved after trimming from an initial average of 30,000,698 \pm 3,715,368 sequences. The percentage of sequences mapped against the reference genome ranged between 37.9 and 45 (Table 1). After filtering, we identified 147,118 SNPs distributed throughout 12,688 complete and partial sequences (SNPs per sequence: mean \pm SD = 11.6 \pm 14.51). Posterior probability per SNP was higher than 0.985 for all SNPs (mean \pm SD = 0.9995 \pm 0.0011). In total, 8023 polymorphic sequences were annotated. Their most common function was related to cellular processes, metabolic processes and biological regulation (Fig. 1, Table S1). Annotated sequences were deposited in the GenBank (BioProject ID: PRJNA528948). This extensive dataset provides a novel genomic resource for *S. ciliata*, and a significant step towards a better understanding of its genetics.

According to identified SNPs, individuals from high and low elevation presented similar values of genetic diversity (Table 1, all Wilcoxon rank tests: $P \geq 0.2$). F_i was positive in five out of six plants (Table 1) and did not differ between elevations (Wilcoxon rank test: $P = 0.7$). Previous studies

on *S. ciliata* using neutral markers also found similar estimates of genetic diversity across elevations (García-Fernández *et al.*, 2012; Lara-Romero *et al.*, 2016). Overall, 775 sequences carrying SNPs associated with elevation were selected by at least one of the implemented approaches, but both shared only 130 of them. Almost 90% ($n = 118$) of them were successfully annotated (Table S1), but they were not significantly enriched in any biological process after FDR correction. However, about 15% of these sequences were associated with responses to stress ($n = 19$) and abiotic stimulus ($n = 16$) (Table S2). This is particularly interesting for the identification of loci related to adaptation to climate change (Giménez-Benavides *et al.*, 2007, 2018), and improving the understanding of adaptation processes in the species.

Supplementary material

The supplementary material for this article can be found at <https://doi.org/10.1017/S1479262119000157>

Acknowledgements

Samples were sequenced in the ETH Zurich Genetic Diversity Centre (GDC). Niklaus Zemp from the GDC for providing access to the reference transcriptome of *Silene latifolia*. Alex Widmer and the Plant Ecological Genetics group (ETH Zurich) for their helpful recommendations and discussions throughout the study. This work was supported by projects AdAptA (CGL2012-33528) and EVA (CGL2016-77377-R) of the Spanish Ministry of Economy and Competitiveness (MINECO). CLR was supported by a Juan de la Cierva post-doctoral fellowship (MINECO: FJCI-2015-24712) and by a European Science Foundation

ESF Exchange Grant (Reference No. 4794) within the ESF activity entitled 'Conservation Genomics: Amalgamation of Conservation Genetics and Ecological and Evolutionary Genomics'. SSB was supported by a FPI pre-doctoral fellowship (MINECO: BES-2017-082317).

References

- Altschul SF, Gish W, Miller W, Myers EW and Lipman DJ (1990) Basic local alignment search tool. *Journal of Molecular Biology* 215: 403–410.
- Bairoch A and Apweiler R (2000) The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Research* 28: 45–48.
- Bernasconi G, Antonovics J, Biere A, Charlesworth D, Delph LF, Filatov D, Giraud T, Hood ME, Marais GAB, McCauley D, Pannell JR, Shykoff JA, Vyskot B, Wolfe LM and Widmer A (2009) *Silene* as a model system in ecology and evolution. *Heredity* 103: 5–14.
- Bolger AM, Lohse M and Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England)* 30: 2114–2120.
- Council Directive 92/43/EEC of 21 May 1992. On the conservation of natural habitats and of wild fauna and flora. *Official Journal of the European Union* 206: 7–50.
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST and McVean G (2011) The variant call format and VCFtools. *Bioinformatics (Oxford, England)* 27: 2156–2158.
- Dray AM (1985) *Plantas a proteger en Portugal continental*. Serviço Nacional de Parques, Reservas e Conservação da Natureza, Lisboa.
- Escudero A, Gimenez-Benavides L, Iriondo JM and Rubio A (2005) Patch dynamics and islands of fertility in a high mountain Mediterranean community. *Arctic, Antarctic and Alpine Research* 36: 518–527.
- Fernández AP, Prieto JAC, Altuna JG, Arregui JL, Gutiérrez LO, Sánchez SP and Janices JV (2007) Flora amenazada presente en la región eurosiberiana de la Comunidad Autónoma del País Vasco. *Naturalia Cantabrigiae* 3: 79–91.
- García-Fernández A, Segarra-Moragues JG, Widmer A, Escudero A and Iriondo JM (2012) Unravelling genetics at the top: mountain islands or isolated belts? *Annals of Botany* 110: 1221–1232.
- García-Fernández A, Escudero A, Lara-Romero C and Iriondo JM (2015) Effects of the duration of cold stratification on early life stages of the Mediterranean alpine plant *Silene ciliata*. *Plant Biology* 17: 344–350.
- Gayral P, Melo-Ferreira J, Glemin S, Bierre N, Carneiro M, Nabholz B, Lourenco JM, Alves PC, Ballenghien M, Faivre N, Belkhir K, Cahais V, Loire E, Bernard A and Galtier N (2013) Reference-free population genomics from next-generation transcriptome data and the vertebrate-invertebrate gap. *PLoS Genetics* 9: e1003457.
- Giménez-Benavides L, Escudero A and Iriondo JM (2007) Local adaptation enhances seedling recruitment along an altitudinal gradient in a high mountain Mediterranean plant. *Annals of Botany* 99: 723–734.
- Giménez-Benavides L, Escudero A, García-Camacho R, García-Fernández A, Iriondo JM, Lara-Romero C and Morente-López J (2018) How does climate change affect regeneration of Mediterranean high-mountain plants? An integration and synthesis of current knowledge. *Plant Biology* 20(Suppl. 1): 50–62.
- Kyrkou I, Iriondo JM and García-Fernández A (2015) A glacial survivor of the alpine Mediterranean region: phylogenetic and phylogeographic insights into *Silene ciliata* Pourr. (Caryophyllaceae). *PeerJ* 3: e1193.
- Lara-Romero C, García-Fernández A, Robledo-Arnuncio JJ, Roumet M, Morente-López J, López-Gil A and Iriondo JM (2016) Individual spatial aggregation correlates with between-population variation in fine-scale genetic structure of *Silene ciliata* (Caryophyllaceae). *Heredity* 116: 417–423.
- Légifrance (2019) 10 mai 1990. Relatif à la liste des espèces végétales protégées en région Auvergne complétant la liste nationale. *Journal Officiel de la République Française*. NOR: PRME9061196A.
- Li H and Durbin R (2010) Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics (Oxford, England)* 26: 589–595.
- López-Moreno JI, Vicente-Serrano JI and Lanjeri S (2007) Mapping snowpack distribution over large areas using GIS and interpolation techniques. *Climate Research* 33: 257–270.
- Marano LA, Marcorin L, Castelli EC and Mendes-Junior CT (2017) Evaluation of MC1R high-throughput nucleotide sequencing data generated by the 1000 Genomes Project. *Genetics and Molecular Biology* 40: 530–539.
- Muller MH, Latreille M and Tollon C (2011) The origin and evolution of a recent agricultural weed: population genetic diversity of weedy populations of sunflower (*Helianthus annuus* L.) in Spain and France. *Evolutionary Applications* 4: 499–514.
- Sanz JMH, Carreño MÁC and Albacete EA (2010) Conservación de flora amenazada en Castilla-La Mancha. *Foresta* 47: 16–28.
- Stölting KN, Paris M, Meier C, Heinze B, Castiglione S, Bartha D and Lexer C (2015) Genome-wide patterns of differentiation and spatially varying selection between postglacial recolonization lineages of *Populus alba* (Salicaceae), a widespread forest tree. *New Phytologist* 207: 723–734.
- Swarts K, Li H, Romero Navarro JA, An D, Romay MC, Hearne S, Acharya C, Glaubitz JC and Mitchell S (2014) Novel methods to optimize genotypic imputation for low-coverage, next-generation sequence data in crop plants. *The Plant Genome* 7: 1–12.
- Turner TL, Bourne EC, Von Wettberg EJ, Hu TT and Nuzhdin SV (2010) Population resequencing reveals local adaptation of *Arabidopsis lyrata* to serpentine soils. *Nature Genetics* 42: 260.
- Ye J, Zhang Y, Cui H, Liu J, Wu Y, Cheng Y, Xu H, Huang X, Li S, Zhou A, Zhang X, Bolund L, Chen Q, Wang J, Yang H, Fang L and Shi C (2018) WEGO 2.0: a web tool for analyzing and plotting GO annotations, 2018 update. *Nucleic Acids Research* 46: W71–W75.