

**Evolution of mating system in diploid sexual types of  
*Cyrtomium falcatum***

July 2016

**Ryosuke IMAI**

Graduate School of Science

CHIBA University

## TABLE OF CONTENTS

GENERAL INTRODUCTION · · · · ·	1
CHAPTER 1 The correlation among mating system, genetic diversity and demography in <i>Cyrtomium falcatum</i> subsp. <i>littorale</i> in Japan · · · · ·	7
CHAPTER 2: Divergence of <i>Cyrtomium falcatum</i> subsp. <i>littorale</i> and the ancestral subsp. <i>australe</i> : phylogenetic inferences from RAD-seq data · · · · ·	47
GENERAL DISCUSSION · · · · ·	71
ACKNOWLEDGEMENTS · · · · ·	75
SUPPORTING INFORMATION · · · · ·	76

## GENERAL INTRODUCTION

The diversity of mating and sexual systems in land plants is of great interest to evolutionary biologists. For seed plants, the presence of both male and female reproductive organs within individuals is a common and ancestral state [1]. Thus, seed plants are potentially faced with a strategic decision on whether to reproduce through outcrossing, selfing, or mixed mating, which is a mixture of outcrossing and selfing [2]. In the short term, selfing is favored due to transmission advantage [3] and reproductive assurance [4,5]. Although the selfed progeny may suffer inbreeding depression [6], selection favoring selfing can operate unless the fitness of outcrossed progeny is more than twice that of selfed progeny [7]. Considering the joint evolution of inbreeding depression and selfing rates, Lande and Schemske (1985) [8] predicted that predominant selfing and predominant outcrossing were alternative stable states of mating system evolution, and mixed mating (intermediate selfing rates) were not stable. In the accompanying meta-analysis, they found that the empirical frequency distribution of outcrossing rates supported their theoretical prediction of a bimodal distribution of outcrossing rates [9]. However, subsequent studies using more species samples [10,11] showed that wind-pollinated species were indeed bimodal, but for animal-pollinated species, mixed mating (selfing rate: 0.2-0.8) was common.

In this respect, the distribution of mating systems in homosporous pteridophytes may be interesting because they are unique among vascular plants; the free-living gametophytes of homosporous pteridophytes can bear both male and female gametangia (antheridium and archegonium) and are capable of three types of mating: intragametophytic selfing (selfing within a gametophyte), intergametophytic selfing (analogous to selfing in seed plants), and intergametophytic crossing (analogous to outcrossing in seed plants) [12]. To empirically test the prediction of a binomial distribution in outcrossing rates in this unique group of plants, Soltis and Soltis (1990) [13] reviewed studies on the mating system of homosporous pteridophytes, and showed that the distribution of outcrossing rates among species was clearly uneven. This result may be important because fertilization is mediated by abiotic factors, such as the spatial dispersal pattern of spores and the water film covering gametophytes, and not by

animals, similar to wind-pollinated seed plants. However, the paucity of evidence for mixed mating in homosporous pteridophytes may simply be due to the lack of studies with a sufficient number of populations. For example, Ranker (2000) [14] examined 23 populations of *Odontosoria chinensis* in the Hawaiian Islands and found that eight of them suggest mixed mating. Furthermore, recent studies using microsatellite markers revealed mixed mating in several fern species [15,16]. Therefore, it is apparent that additional empirical studies on mating system are required to more comprehensively determine the nature of mating system distribution in homosporous pteridophytes.

In addition to the taxonomic distribution of mating system, changes in population genetic features along with mating system transition have not been well examined in homosporous pteridophytes. For example, the correlation between the evolution of selfing and severe genetic drift events such as bottlenecks has been considered to be important because both the transmission advantage [3] and reproductive assurance [4,5] theories predict their co-occurrence. The former theory predicts that population bottleneck events that result in the purging of inbreeding depression would trigger the evolution of selfing [8], while in the latter theory, selfing would be favoured under limited mating opportunities, which are expected in colonization processes. Thus, if populations display mating system variation, reduction of effective population size is expected, due to the co-occurrence of selfing and bottlenecks [8]. However, this scenario has not been well examined in this unique plant groups.

In this thesis, I used sexual diploid cytotype of *Cyrtomium falcatum* (L. f.) C. Presl (Dryopteridaceae). Three cytotypes have been identified in *C. falcatum*; sexual diploid, apogamous triploid, and sexual tetraploid [17, 18, 19]. Matsumoto (2003) [20] divided the sexual diploid race in Japan into two subspecies, subsp. *littorale* and subsp. *australe*. The two subspecies occupy nearly parapatric distribution in Japan: subsp. *littorale* from the southern part of Hokkaido Island to the Shikoku Island, and subsp. *australe* in the Ogasawara Islands and from Kyushu to the Nansei Islands (Fig. 1 in Chapter 1). Subspecies *littorale* is distinguished from subsp. *australe* by smaller blades, fewer pairs of pinnae, and grayish indusia without a blackish brown center. Subspecies *littorale* grows on the coastal rocks or cliffs. Subspecies *australe* grows mainly on the floor of maritime forest, but is found

occasionally on the coastal rocks or cliffs just as subsp. *littorale*.

The present thesis consists of two chapters. In the first chapter, mating system divergence among populations in *C. falcatum* subsp. *littorale* was examined by using microsatellite markers. The selfing rates were shown to be correlated significantly with the gametangium formation types (the S-type and the M-type: two sexual expression types on gametophytes). Additionally, the Approximate Bayesian Computation (ABC) analyses suggested that effective population sizes ( $N_e$ ) of the high selfing populations (the M-type ones) were about a third of that of the S-type populations. These results suggest that the M-type populations have experienced more frequent bottlenecks, which could be related to their higher colonization ability via intragametophytic selfing. I discuss on the evolutionary factors that had caused the mating system divergence among populations in this subspecies.

In the next chapter, phylogenetic relationship among populations of both subspecies and demographic changes along lineages were examined by using huge Single Nucleotide Polymorphisms (SNPs) data generated by Next Generation Sequencer. Matsumoto (2003) suggested that *C. falcatum* subsp. *australe* is an outcrosser based on its S-type of sexual expression of gametophytes, and the consistently low rates of sporophyte formation in isolated cultures. Because of this intraspecific mating system variation, two sexual diploid subspecies of *C. falcatum* provide an interesting experimental opportunity to examine the evolutionary pattern of mating system transition. The phylogenetic analysis revealed three distinct groups: 1) LITTORALE (subsp. *littorale* except SADO and some of SAND), 2) KYUSHU (mainly Kyusyu populations of subspecies *australe*) and 3) OGASAWARA (Ogasawara populations of subspecies *australe*). Among the subsp. *littorale* populations, all populations except SADO showed significantly positive  $F_{IS}$  values, suggesting that they were mixed mating populations (chap. 1). Because the populations with positive  $F_{IS}$  values formed a well-supported group of LITTORALE, mixed mating could have originated only once from the ancestral state of outcrossing in diploid race of *C. falcatum*.

Lastly, we discussed importance of genetic load on evolution of mixed mating of *C. falcatum*. Low genetic load could be caused by population bottleneck under speciation. That might create mating system difference between two subspecies.

## REFERENCE

1. Yampolsky C YH. Distribution of sex forms in the phanerogamic flora. *Bibl Genet.* 1922;3.
2. Busch JW, Delph LF. The relative importance of reproductive assurance and automatic selection as hypotheses for the evolution of self-fertilization. *Ann Bot.* 2012;109: 553–62.
3. Fisher R a. Average excess and average effect of a gene substitution. *Annu Eugen.* 1941;11: 53–63.
4. Baker HG. Self-Compatibility and Establishment After “Long-Distance” Dispersal. *Evolution.* 1955;9: 347–9.
5. Morgan MT, Wilson WG. Self-fertilization and the escape from pollen limitation in variable pollination environments. *Evolution.* 2005;59: 1143–8.
6. Darwin C. The effects of cross and self fertilisation in the vegetable kingdom. J. Murray; 1876.
7. Lloyd DG. Some Reproductive Factors Affecting the Selection of Self-Fertilization in Plants. *Am Nat.* 1979;113: 67–79.
8. Lande R, Schemske DW. the Evolution of Self-Fertilization and Inbreeding Depression in Plants .1. Genetic Models. *Evolution.* 1985;39: 24–40.
9. Schemske DW, Lande R. The evolution of self-fertilization and inbreeding depression in plants. II. Empirical observations. *Evolution.* 1985;39: 41–52.

10. Vogler DW, Kalisz S. Sex among the flowers: the distribution of plant mating systems. *Evolution*. 2001;55: 202–4.
11. Goodwillie C, Kalisz S, Eckert CG. THE EVOLUTIONARY ENIGMA OF MIXED MATING SYSTEMS IN PLANTS: Occurrence, Theoretical Explanations, and Empirical Evidence. *Annu Rev Ecol Evol Syst*. 2005;36: 47–79.
12. Klekowski EJ. Reproductive biology of the Pteridophyta. II. Theoretical considerations. *Bot J Linn Soc*. 1969;62: 347–59.
13. Soltis PS, Soltis DE. Evolution of Inbreeding and Outcrossing in Ferns and Fern-Allies. *Plant Species Biol*. 1990;5: 1–11.
14. Ranker T a, Gemmill CE, Trapp PG. Microevolutionary patterns and processes of the Native Hawaiian colonizing fern *Odontosoria chinensis* (Lindsaeaceae). *Evolution*. 2000;54: 828–39.
15. De Groot G a., During HJ, Ansell SW, Schneider H, Bremer P, Wubs ERJ, et al. Diverse spore rains and limited local exchange shape fern genetic diversity in a recently created habitat colonized by long-distance dispersal. *Ann Bot*. 2012;109: 965-78.
16. Chung MY, Moon MO, López-Pujol J, Maki M, Yamashiro T, Yukawa T, et al. Was Jeju Island a glacial refugium for East Asian warm-temperate

- plants? insights from the homosporous fern *Selliguea hastata*  
(Polypodiaceae). *Am J Bot.* 2013;100: 2240–9.
17. Manton I, others. Problems of cytology and evolution in the Pteridophyta. *Probl Cytol Evol Pteridophyta.* 1950;
  18. Matsumoto S. Chromosome numbers and the reproductive modes of Japanese *Cyrtomium*. *Ann Tsukuba Bot Gard.* 1985;3: 1-7.
  19. Lu JM, Cheng X, Wu D, Li DZ. Chromosome study of the fern genus *Cyrtomium* (Dryopteridaceae). *Bot J Linn Soc.* 2006;150: 221-8.
  20. Matsumoto S. Species ecological study on reproductive systems and speciation of *Cyrtomium falcatum* complex (Dryopteridaceae) in Japanese archipelago. *Ann Tsukuba Bot Gard.* 2003;22: 1-141.

## **CHAPTER 1 The correlation among mating system, genetic diversity and demography in *Cyrtomium falcatum* subsp. *littorale* in Japan.**

### **INTRODUCTION**

In chapter 1, we focused on Japanese holy fern, *Cyrtomium falcatum* subsp. *littorale*, to evaluate the evolutionary relationship among sexual expression of gametophytes, mating system, genetic diversity, and demographic history. Three cytotypes have been identified in *C. falcatum*; sexual diploid, apogamous triploid, and sexual tetraploid [1–3]. From these three cytotypes, we used northern type of diploid, *C. falcatum* subsp. *littorale* S.Matsumoto nom. nud. (Dryopteridaceae). Matsumoto (2003) [4] divided the sexual diploid *C. falcatum* in Japan into two subspecies, subsp. *littorale* and subsp. *australe*. Subspecies *littorale* is distinguished from subsp. *australe* by smaller blades, fewer pairs of pinnae, and grayish indusia without a blackish brown center. Subspecies *littorale* grows on coastal rocks or cliffs and is distributed discontinuously from the southern part of Hokkaido Island to Shikoku Island ([4]; Fig. 1.1). Matsumoto (2003) [4] observed variation in sexual expression in the northern type of *C. falcatum*. Sporophytes called mixed type (M-type) simultaneously produce gametophytes with both antheridia (male gametangia) and archegonia (female gametangia) at frequencies of 90% and over at three months in cultivation. In contrast, sporophytes of the separate type (S-type) produce these bisexual gametophytes at frequencies of 10% and less at the same stage. This sexual expression is equivalent to dichogamy, because the gametophytes produce antheridia in early stages of development and then produce archegonia after releasing sperm. The sporophyte produces intermediate frequency ( $10\% < \text{frequency} < 90\%$ ) of bisexual gametophytes was considered as intermediate type (I-type). Traditionally, the S-type of gametangium formation on gametophytes (femaleness at maturity) has been considered a

morphological adaptation to promote intergametophytic mating and the M-type (bisexuality at maturity) an adaptation to intragametophytic selfing [5,6]. In support of this hypothesis, gametophytes of the M-type sporophyte can produce sporophytes at very high rates (84%–100%) in isolated cultures, whereas S-type sporophytes show a low rate of sporophyte formation (5%–35%) [4]. The geographical distribution pattern of the three gametangium formation types [4] suggests that there is mating system population differentiation in this subspecies. Understanding this variation is important because it may provide clues to the causes of mating system evolution. For example, a series of studies on tristylous *Eichhornia paniculata* indicated that highly selfing populations were derived independently from semi-homostylous morphs [7–9]. In addition, they found, as theoretical studies predicted, that genetic diversity and effective population size were reduced, while inbreeding coefficients increased in selfing populations. The correlation between the evolution of selfing and population bottleneck seems to be important because both transmission advantage[10] and reproductive assurance[11] theories predict their co-occurrence. In order to know how mating systems differentiate at population level, therefore, inference of demographic history of each population would be critical. Recent advances in coalescent-based analysis in population genetics have made it possible to infer demographic history and address questions about the evolution of selfing [12,13]. Particularly, Approximate Bayesian Computation (ABC) provide flexible approach in the demographic inference, allowing many questions relevant to ecology and evolutionary biology to be addressed [14]. In the present study, we developed eight new microsatellite markers to examine genetic variations among populations of *C. falcatum* subsp. *littorale*. By analyzing these markers, we addressed the following questions. 1) Do the wild populations of M-type individuals show higher selfing rates than those of the S-type individuals? 2) If the selfing rate varies among populations, are there correlations between high selfing rates, low genetic diversity, and historical population bottleneck events? 3) Is the M-S-types

variation in mating system reflected in range-wide genetic structure?

## MATERIALS AND METHODS

### *Populations and sampling*

Matsumoto (2003)[4] examined the gametangium formation types of 33 individuals of *C. falcatum* subsp. *littorale* from 20 localities (Fig. 1.1). In this study, we collected 233 samples from seven populations (21–42 individuals per population) in five localities. The seven populations and their localities were as follows: ESAN 1 (41.8112N, 141.1844E) and ESAN 2 (41.8115N, 141.1844E) from Esan-misaki, Hokkaido Prefecture; SADO (38.0929N, 138.2498E) from Sado Island, Niigata Prefecture; IZU 1 (34.8824N, 139.1323E) and IZU 2 (34.8821N, 139.1319E) from Jogasaki, Shizuoka Prefecture; KANT (33.6004N, 135.6004E) from Kantori-misaki, Wakayama Prefecture; and SAND (33.6655N, 135.3355E) from Sandan-peki, Wakayama Prefecture (Fig. 1.1). Three populations (IZU 1, IZU 2, SADO) were selected from the localities where S-types have been observed, and four populations (ESAN 1, ESAN 2, SAND, KANT) were collected from the localities where M-types have been observed [4]. Voucher specimens of the samples (ri010001–ri010233) were deposited in the Herbarium of the National Museum of Nature and Science (TNS), Tsukuba, Ibaraki, Japan.

### *DNA extraction and microsatellite marker development*

Total DNA was extracted from silica-dried leaves using the HEPES/CTAB method [15]. The DNA samples were used for microsatellite marker development and further genotyping. We developed microsatellite markers using two different methods; one was an improved technique for isolating co-dominant compound microsatellite markers [16] and the other was a next-generation sequencing (NGS) method [17].

Firstly, following the method of Lian *et al.* (2006) [16], genomic DNA from a sample of IZU1 was digested with six blunt-end cutters (*Hae*III, *Pvu*II, *Alu*I, *Ssp*I, *Eco*RI, and *Sca*I) and ligated with a specific blunt adaptor [18] using a T4 DNA Ligation kit (Nippon Gene, Tokyo, Japan). The digested and ligated fragments were

amplified using a compound SSR primer (AC)<sub>5</sub>(AG)<sub>8</sub> and an adaptor primer AP2 (5'-CTATAGGGCACGCGTGGT-3') [16]. The PCR products were cloned using the TOPO-TA Cloning Kit (Invitrogen, Carlsbad, USA). Plasmid DNAs were amplified from the colonies with a TempliPhi DNA Amplification Kit (GE Healthcare Bio-Sciences, Little Chalfont, UK). Sequence reactions were prepared with T3 and T7 primers (Invitrogen) using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Tokyo, Japan). The reaction mixture was analyzed on an ABI 3500 genetic analyzer (Applied Biosystems). A total of 576 different fragments with a compound microsatellite motif at one end were obtained. Specific primers were designed from 86 different sequences with PRIMER3 software [19]. A PIG1-tail (5'-GTTTCTT-3') was added to specific forward primers to reduce stuttering [20].

Secondly, we developed microsatellite markers using an NGS method with a Roche 454 Genome Sequencer Junior (Roche/454 Life Sciences, Branford, CT, USA). Genomic DNA was isolated from a pinna of *Cyrtomium falcatum* subsp. *littorale*, collected from the Tsukuba Botanical Gardens (originally collected from Wakayama Pref., Japan) and fragmented by nebulization. A DNA library was constructed using the GS FLX Titanium Rapid Library Preparation Kit (Roche/454 Life Sciences). The DNA library was purified using the MinElute PCR Purification Kit (Qiagen, Tokyo, Japan) and its quality was checked using the Agilent High Sensitivity DNA kit (Agilent Technologies, Palo Alto, CA, USA). Emulsion PCR was carried out using the GS Junior Titanium emPCR Lib-L Kit (Roche/454 Life Sciences), and pyrosequencing was conducted on a Roche 454 Genome Sequencer Junior instrument at the Tsukuba Botanical Gardens, with the GS Junior Titanium Sequencing Kit (Roche/454 Life Sciences). Contigs were assembled to over 500bp with GS Newbler De Novo Assembler (Roche/454 Life Sciences), implementing the default parameters and heterozygotic mode. The program QDD v.2.1 [21] was used with default settings to detect and select microsatellite sequences. Twelve hundred contigs were used for

searching microsatellite candidates. We designed 72 primer pairs based on the penalty scores calculated with Primer3 in the QDD pipeline. As described by Schuelke (2000) [22], the U19 sequence (5' -GGTTTCCAGTCACGACG-3' ) was added to the 5' end of specific forward primer sequences. A PIG2 tail (5' -GTTT-3' ) was added to specific reverse primer sequences [20].

#### *Fragment analysis of microsatellite markers*

We tested all of the candidate markers (86 by the method of Lian *et al.* (2006) and 72 by the NGS method) for good PCR amplification, reproducibility, and the level of polymorphism, using a subset of samples: two individuals from each of the seven populations. Finally, eight primer pairs were selected and used to further genotype all samples (Table 1.1). PCR amplifications (simplex PCR) were performed using the Multiplex PCR Kit (Qiagen) in a downscaled final volume of 5 µl according to the manufacturer's protocol. For two primer sets, CFL-079 and CFL-C32, PCRs were conducted using each specific primer and a dye-labeled compound microsatellite primer under the following conditions: initial denaturation for 15 min at 95°C, followed by 30 cycles at 95°C for 30 s, 55°C for 90 s, 72°C for 1 min, and final extension at 60°C for 30 min. The PCR reaction mixture for the other six primer sets contained 0.2 µM reverse primer, 0.2 µM fluorescent dye-labeled U19 primer (ABI PRISM®, Applied Biosystems), and 0.04 µM forward primer. Touchdown PCR was performed with initial denaturation for 15 min at 95°C, followed by 25 cycles at 95°C for 30 s, 63–53°C (with a 0.5°C decrease for every subsequent cycle) for 90 s, and 72°C for 1 min, followed by 20 cycles of 95°C for 30 s, 53°C for 90 s, and 72°C for 1 min, and final extension at 60°C for 30 min. The PCR products were analyzed on an ABI 3500 Genetic Analyzer (Applied Biosystems) with the internal size standard, GeneScan 600 LIZ (Applied Biosystems), and fragment sizes were determined with GeneMapper 3.1 (Applied Biosystems). The original sequences for the markers were deposited in GenBank under

the accession numbers LC055975 - LC055982 (Table 1.1).

### *Data analyses*

#### *Inbreeding coefficient and genetic diversity within populations*

Gene diversity ( $h$ ; [23]), allelic richness ( $A_R$ ; [24]), and the inbreeding coefficients ( $F_{IS}$ ; [25]) were calculated for each population in FSTAT ver. 2.9.3.2 (hereafter, FSTAT, [26]). FSTAT was also used to test genotypic disequilibrium among loci for each population. We used INEST2 [27], with the ‘nfb’ model, to estimate  $F_{IS}$  values within populations, taking into account the effect of underestimating heterozygosity in the presence of null alleles. As mentioned previously, two types of self-fertilization may occur in homosporous ferns: intragametophytic ( $S_I$ ) and intergametophytic ( $S$ ) selfing. Hedrick (1987) [28] derived a formula showing the relationship among  $F_{IS}$  and the rates of two types of selfing ( $S_I$  and  $S$ ).

$$F_{IS} = \frac{S + 2S_I}{2 - S}$$

If we expect no intergametophytic selfing ( $S = 0$ ),  $F_{IS}$  equals  $S_I$  as shown by McCauley et al. (1985) [29]. In the present study, we use  $F_{IS}$  values as an indicator of the relative contributions of selfing *sensu-lato*, and outcrossing [30]. To assess whether population genetic parameters differ between M- and S-type populations, inbreeding coefficients ( $F_{IS}$ ), gene diversity ( $h$ ), allelic richness ( $A_R$ ), relatedness [31], and  $F_{ST}$  [32] values were compared, treating M and S types as two groups. Differences in these values between the two types were tested for significance using a permutation test in FSTAT. We employed one-sided  $P$ -values to test whether the value in one group is significantly larger than the other. Although the distributions of *C. falcatum* subsp. *littorale* and subsp. *australe* have been considered parapatric (Fig. 1.1), our STRUCTURE analysis indicated that it was likely subsp. *australe* was inadvertently included in, and admixed with, our samples, especially in the SAND and SADO populations (see Results and

Discussion for details). Therefore prior to evaluating genetic diversity and demographic inferences, we removed individuals which had ancestry values of greater than 50% to the cluster corresponding to *C. falcatum* subsp. *austorale* in  $K = 7$ .

#### *Genetic differentiation and structure among populations*

Genetic differentiation among populations was evaluated by calculating the overall and pairwise  $F_{ST}$  [32] values and their respective confidence intervals (CI) (95 and 99%) were determined on the basis of 1000 bootstrapping replicates using FSTAT. The standardized values of  $F_{ST}$ ,  $F'_{ST}$  [33] were also calculated using GenAlEx 6.5 [34]. Patterns of isolation by distance (IBD; [35]) were evaluated, using GenAlEx ver. 6.5 [34], according to the method described by Rousset (1997)[36]; a Mantel test with 999 random permutations between the matrices obtained for pairwise population differentiation in terms of  $F_{ST} / (1 - F_{ST})$  and the natural logarithms of direct minimum geographic distance among populations. Genetic structure was also investigated with the model-based clustering algorithm implemented in the software STRUCTURE v. 2.3.3 [37,38]. A number of clusters ( $K$ ) varying from 1 to 15, was evaluated under the correlated allele frequencies model by running 100,000 burn-in Markov Chain Monte Carlo (MCMC) repetitions and 1,000,000 subsequent repetitions based on the LOCPRIOR model [38]. The probabilities of each  $K$  were averaged over 10 runs. We employed the CLUMPAK server (ref, <http://clumpak.tau.ac.il/index.html>) to evaluate multimodality [39] among runs at each  $K$ . The optimum  $K$  value was determined based on  $\Delta K$  [40], evaluating the probability of the data ( $\ln P(D)$ ) for each  $K$  value using STRUCTURE HARVESTER [41]. Bar charts representing the proportion of cluster membership in each individual were obtained using CLUMPAK. The genetic relationships between the clusters were evaluated based on genetic distance calculated in STRUCTURE and a neighbor-joining tree of clusters was generated using Populations 1.2.23 [42].

### *Inference of demographic history*

The software DIYABC v2.0 [43,44] was used to infer the demographic history of *Cyrtomium falcatum* based on the Approximate Bayesian Computation (ABC) approach. DIYABC provides flexibility for the mutation models of microsatellite loci in coalescent simulations, allowing both the generalized stepwise mutation model (GSM; [45]) and the single nucleotide indel model (SNI). As our main purpose was to test whether the effective population size and demographic history among M- and S-type populations were different due to their different gametangium formation, three simple scenarios were examined in each population (Fig. 1.2):

Scenario 1. Bottleneck model: the ancestral effective population size ( $N_a$ ) was changed at  $t_1$  to the modern effective population size ( $N_1$ ) and  $N_1$  was set to be smaller than  $N_a$ .

Scenario 2. Constant model: the ancestral effective population size ( $N_b$ ) and the modern one ( $N_1$ ) were set to be equal, assuming the effective population size has not changed.

Scenario 3. Expansion model: the ancestral effective population size ( $N_c$ ) was changed at  $t_1$  to the modern effective population size ( $N_1$ ) and  $N_1$  was set to be larger than  $N_c$ .

In these scenarios,  $t_1$  represents time scale measured by generation time. We employed the default values of the priors for each parameter in DIYABC. The mean values for expected heterozygosity ( $H_E$ ), number of alleles ( $A$ ), allele size variance across loci and M index across loci [46,47] were used as summary statistics. A million simulations were run for each scenario. After all the simulations had been ran, the most-likely scenario was determined by comparing the posterior probabilities using the logistic regression method. The goodness of fit of the scenario was assessed by the option “model checking” with principal component analysis (PCA) in DIYABC, which measures the discrepancy between the model and real data. To translate the inferred number of

generations for  $t_1$  to time scale by year, we assumed a generation time of 3 years, as Matsumoto (2003) [4] showed that *C. falcatum* subsp. *littorale* produces spores at 1 year in cultivation tests. Although this is a simple ABC test, Sakaguchi et al. (2013) [48] employed a similar approach to successfully detect geographic patterns in population demographic history of conifer species in Australia, therefore we believe this approach could also be informative in our study.

## RESULTS

### *Characteristics of the eight microsatellite loci*

Eight new microsatellite markers were developed in this study (Table 1.1). Amplification size ranges and the number of alleles per loci are shown in Supplementary Data Table 1.S1. The null allele frequencies estimated by INEST2 (Supplementary Data Table 1.S2) were relatively high (over 0.10) in 9 out of 56 (8 locus  $\times$  7 population) combinations, but significant in only one case: CFL-B12 locus in SADO (0.246; 95% CI: 0.0977–0.400). No significant deviations from genotypic equilibrium were detected once putative admixed individuals in SAND and SADO were excluded.

### *Mating system and genetic diversity*

All of the populations except for SADO1 showed significantly positive inbreeding coefficient ( $F_{IS}$ ) values, ranging from 0.220 to 0.794 over all loci (Table 1.2). The average  $F_{IS}$  value of M-type populations (0.632) was significantly higher than S-type (0.153,  $P < 0.05$ ; Tables 1.2, 1.3). The  $F_{IS}$  values estimated using INEST2 were generally lower than those estimated using FSTAT because of the possible presence of null alleles (Table 1.4). Nevertheless,  $F_{IS}$  values were still significantly larger than zero in all populations, except for SADO (Table 1.4), indicating significant levels of selfing. The average allelic richness ( $A_R$ ) of the M-type populations (1.91) was significantly lower than that of the S-type populations (2.61,  $P < 0.05$ ; Tables 1.2, 1.3). Similarly, the average value of gene diversity ( $h$ ) of the M-type populations (0.170) was significantly lower than that of the S-type populations (0.380,  $P < 0.05$ ; Tables 1.2, 1.3).

### *Population genetic structure*

The overall  $F_{ST}$  and  $F'_{ST}$  values were 0.581 and 0.739, respectively, indicating a high level of genetic differentiation among populations. The average of pairwise  $F_{ST}$  values

among the M-type populations was higher than that among the S-type populations, and the difference was nearly significant ( $P = 0.067$ ). Significant IBD ( $R^2 = 0.3447$ ;  $P < 0.05$ ) was detected among the 7 populations (Fig. 1.3). In the STRUCTURE analysis, the mean probability of the data ( $\text{LnP(D)}$ ) increased steadily up to  $K = 7$  (Fig. 1.4) and  $\Delta K$  suggested  $K = 7$  as optimal (Fig. 1.4). At  $K = 2$ , ESAN 1 and 2 were grouped into cluster 1 and the remaining populations were assigned to cluster 2 (Fig. 1.5). Thus, the clustering at  $K = 2$  did not correspond to the M- and S-types. At  $K = 3$ , two M-type populations (KANT and SAND) in Wakayama Prefecture and one S-type population (SADO) were separated from IZU1 and IZU2. At  $K = 4$ , KANT was differentiated. At  $K = 5$  and greater, five clusters corresponding to five main sampling localities (Fig. 1.1) were observed. In  $K = 6$ , SADO and SAND populations were shown to contain a considerable number of admixed individuals (with cluster 6). At  $K = 7$ , cluster 6 was further divided into two clusters and cluster 7 corresponded to the admixed cluster in SAND. The NJ tree for the seven clusters revealed two groups. In one group, cluster 1 (ESAN 1 and 2), 2 (IZU 1 and 2), 4 (KANT) and 5 (major cluster in SAND) were grouped together, while cluster 3 and 6 in SADO and cluster 7 for the admixed cluster in SAND were in the other group. The  $F$  values of each cluster (analogous to the  $F_{ST}$  values between each cluster and the assumed ancestral population) showed that clusters corresponding to M-type populations had larger values (0.627 - 0.751) than S-type populations (0.474 - 0.508, Fig. 1.\*).

#### *Inference of demographic history of each population*

In DIYABC, the highest posterior probability was for scenario 1 (bottleneck model), and its 95% CI did not overlap with those of the other two scenarios in each of the 7 populations, regardless of gametangium formation type (Table 1.S3). For scenario 1, the median values of the effective modern population size of  $N_1$  were well estimated in each population. The S-type populations had significantly larger  $N_1$  values (465 – 846) than

the M-type ones (169 – 328; Table 1.2, Fig. 1.S1;  $t$ -test,  $P < 0.05$ ). However, the posterior distribution pattern suggested that other parameters were poorly estimated (Fig. 1.S1, Table 1.S4), with the exception of the timing of the population size change event ( $t_1$ ) in the SADO population. In the SADO population, the median value of  $t_1$  was 2,940 generations ago (95% CI, 256 – 9,390), corresponding to 8,820 years ago (95% CI, 768 – 28,170). In all populations, all of the summary statistics showed no significant differences between the observed and simulated data, based on the posterior distributions (Table 1.S4), and the PCA showed that the observed data point was centered on the cluster of simulated data points, based on the posterior distributions (Fig. 1.S2), suggesting that scenario 1 was a good fit to the observed data in all populations.

## DISCUSSION

### *Mixed mating to outcrossing in C. falcatum subsp. littorale*

Examples of mixed mating, shown not by isolated gametophyte tests [49] but by genetic markers, are not common in homosporous ferns, but have been reported for some populations of the following species: *Dryopteris expansa* [50], *Blechnum spicant* [51], *Hemionitis palmata* [52], *Odontosoria chinense* [53], *Asplenium scolopendrium* [54] and *Selliguea hastata* [55]. Here, we found evidence for mixed mating in *C. falcatum* subsp. *littorale*. as all but one population (SADO) had significantly positive  $F_{IS}$  values, ranging from 0.2 to 0.8, indicating it is highly likely that both gametophytic selfing and crossing are widespread.

In contrast to our results, Chung et al. (2012) [56] examined sexual *C. falcatum* populations along the southern shores of South Korea, and showed that the  $F_{IS}$  values for these populations did not significantly deviate from zero. The *C. falcatum* individuals described in Chung et al. (2012) may be *C. falcatum* subsp. *australe*, because Matsumoto (2003) [4] reported the distribution of this subspecies in Nagasaki Pref., Japan (Fig. 1.1), which is only 200 km away from southern South Korea, separated by the Tsushima Strait. Matsumoto (2003) [4] suggested that *C. falcatum* subsp. *australe* is an outcrosser based on its S-type of sexual expression of gametophytes, and the consistently low rates of sporophyte formation in isolated cultures. Because of the evolution of these different mating systems, the two sexual diploid subspecies of *C. falcatum* provide an interesting experimental opportunity for future studies.

A interesting finding from this study was the detection of a cluster specific to the SADO and SAND populations in the STRUCTURE analysis ( $K = 6$ , cluster 6, Fig. 1.5). The most likely explanation for this cluster is the sympatric distribution of, and admixture with, *C. falcatum* subsp. *australe*. Although Matsumoto (2003) [4] recognized two sexual diploid forms (subsp. *littorale* and subsp. *australe*) distributed

parapatrically in Japan (Fig. 1.1), the morphological diagnostic characters were limited to frond size and indusium color. Furthermore, artificial crossing experiments showed that hybrids are fertile. Therefore, putative hybrids may be difficult to identify morphologically, particularly backcross hybrids. These backcross could have some influence in mating system or morphological characters. Additional studies including samples of subsp. *australe* would be required to clarify the geographical distribution pattern of the two subspecies of *C. falcatum*.

#### *Genetic diversity and mating system*

The  $F_{IS}$  values of M-type populations were significantly higher than those of S-type ones (Table 1.3). Allelic richness ( $A_R$ ) and gene diversity ( $h$ ) values of M-type populations were lower than those of S-type ones and the effective population size estimated by DIYABC also showed the same pattern (Tables 1.2 and 3). To our knowledge, this is the first study to show that the gametangium formation types of gametophytes affect the levels of inbreeding and genetic diversity in natural populations of homosporous ferns. It is well known that inbreeding species have lower neutral genetic diversity within populations compared to more outcrossing taxa [57]. The correlation between genetic diversity and gametangium formation M- and S-types in *C. falcatum* subsp. *littorale* follows these general trends. The observed patterns could be due to several factors. Firstly, inbreeding is expected to reduce effective population size:  $N_e = N / (1+F_{IS})$  [58]. Secondly, recent empirical studies in seed plants have revealed that reduction of genetic diversity or effective population size are often greater than those expected from  $F_{IS}$  values alone [59], possibly because of linked selection owing to reduced recombination efficiency [60], and/or because of population bottlenecks.

In the present study, despite their intermediate  $F_{IS}$ , average gene diversity ( $h$ ) of the M-type populations (0.152) was about half that of S-type populations (0.367), and the average  $N_e$  of the M-type populations (215) was about a third of that of S-type

populations (675). Although the 95% CIs of the inferred  $N_e$  values in the ABC should be taken into consideration, these levels of reduction in  $h$  and  $N_e$  are comparable to the case of complete inbreeding, and seem to be greater than those expected from the intermediate  $F_{IS}$  values observed in M- and S-type populations (0.626 vs. 0.208). Intragametophytic selfing in ferns has been thought to be an advantage for colonization to a distant place as it enables a single spore to establish a new population [36,49,50]. For example, Groot et al. (2012) [54] examined fern populations in a recently reclaimed Dutch polder land and concluded that the polder land was colonized via multiple independent single-spore colonization events in all four species studied. It is likely that simultaneous formation of both male and female gametangia and higher rates of selfing confer higher colonization ability to M-type individuals. Both M- and S-type individuals of *C. falcatum* subsp. *littorale* are lithophytes that grow on sea cliffs, a habitat that is vulnerable to disasters such as a landslides. In fact, one population of *C. falcatum* subsp. *littorale* in Fukushima Pref. was lost after the 2011 Great East Japan Earthquake and Tsunami [63]. DIYABC detected reduction of population size (Scenario 1, Fig. 1.2) in all populations regardless of M- or S-type (Table 1.S3), and this might reflect past episodes of local extinction and recolonization in unstable habitats. Although there is no apparent difference between the habitats of M- and S-types, it is possible that differences in magnitude or frequency of past colonization bottleneck events could result in significant differences in the genetic diversity and effective population sizes between M- and S-type populations.

#### *Population genetic structure*

Previous studies using allozymes or microsatellites showed that  $F_{ST}$  values are highly variable in homosporous pteridophyte species as in seed plant species:  $F_{ST} = 0.065$  for *Odontosoria chinensis* in Hawaii [53],  $F_{ST} = 0.520$  for *Dryopteris aemura* in Spain [64],  $F_{ST} = 0.543$  for *Cyrtomium falcatum* in Korea [56], and  $F_{ST} = 0.390$  for *Selliguea*

*hastate* in Japan [55]. The  $F_{ST}$  and  $F'_{ST}$  values (0.581 and 0.739, respectively) of the present study are comparatively high. Interestingly, Korean outcrossing populations of *C. falcatum* [56] have a similar overall  $F_{ST}$  value (0.543) to *C. falcatum* subsp. *littolare*, indicating this level of population differentiation is typical for the species and could be due to species dispersal ability and/or habitat preferences. For example, it is possible that discontinuous geographical distribution of the species' suitable habitat, such as crevices in sea cliffs or rocks near the seashore, is responsible for population differentiation. Significant IBD was detected in this study (Fig. 1.3) and the STRUCTURE analysis did not show a clear clustering corresponding to the M- and S-type populations (Fig. 1.5). Although more populations would be required to conclude, we suggest that range-wide genetic structure of this species is not generated by M- and S-type differences but rather by geographic distance.

Significant IBD, related to past range shifts following climate change, has been detected in many plant species in the Japanese archipelago [65–67]. Significant IBD was also reported in a fern species, *Asplenium fontanum* subsp. *fontanum*, in Europe [68] and was likely caused by range expansion following the last glacial maximum (LGM). For fern species in Japan, the past distributional shifts in relation to the LGM are not well examined in palaeoecological studies of spore fossils. However, pollen fossil studies [69] suggest there were refugia areas for warm temperate tree species during the LGM on the Izu and Kii peninsulas, which were examined in this study. Thus, it is reasonable to consider that the genetic structure detected in this study was influenced by past distributional shifts from these refugia. Interestingly, although the time scale of population size change in *C. falcatum* subsp. *littorale* was well estimated in only one population (SADO, 8820 years ago), this timing corresponds to a post-LGM recolonization scenario. Moreover, STRUCTURE analysis revealed seven clusters, suggesting multiple lineages related to past colonization episodes. In particular, the ESAN population is located near the northern range-limit of *C. falcatum* in Hokkaido

and revealed a cluster specific to Hokkaido at  $K = 2$  in the STRUCTURE analysis. This result suggests the persistence of *C. falcatum* during the LGM in an area further north than the well-defined refugia (e.g. Izu and Kii peninsulas), as often discussed in cold-tolerant plant species [66,70–72]. Sato and Sakai (1978)[73] reviewed the distributional ranges of 594 fern species in Japan and categorized 10 groups based on their northern range-limit patterns. According to their study, *C. falcatum* was categorized into a group whose range-limit is in Hokkaido, more northerly than the other 9 groups. Therefore, *C. falcatum* may have some degree of cold-tolerance. Indeed, although species were not distinguished, palaeoecological studies detected fern spores in Hokkaido even during the LGM and their expansion started from around 8000 years ago [74]. Thus, although *C. falcatum* may have been mainly distributed in southern refugia on Honshu Island during the LGM, its persistence in Hokkaido or the northern part of Honshu is also a possibility, as discussed in Tsuda et al. (2015)[72] and Kitamura et al. (2015)[75].

The M- and S-types that were defined based on sexual expression patterns of gametophytes were significantly different in levels of selfing ( $F_{IS}$ ), genetic diversity, and effective population size. These results suggest that reproductive and demographic differences exist between the two types, despite the lack of genetic structure between them. The more severe population bottlenecks were inferred to have 497 occurred in the M-type populations rather than the S-types. This implies that extinction and recolonization events can provide opportunities for the maintenance, and possibly evolution of the M-type sexual expression in each population. Evolution of selfing via transmission advantage operates only under a low genetic load, and fixed selfing is expected [11]. Therefore, the existence of mixed mating even in the M-type populations would suggest that reproductive assurance, rather than transmission advantage, is the main factor affecting the evolution of selfing in this species. If this is the case, the M-type would be advantageous in small and disturbed populations, while the S-type is

advantageous in large and stable populations.

In this study, we found that the M- and S-type variation in mating system is highly related to genetic diversity within populations but not range-wide genetic structure. This indicates that historical events at a wider spatio-temporal scale (e.g. post-LGM recolonization) had more impact on genetic structure. The inference of range wide demographic history (e.g. time scale of divergence, admixture, the amount and direction of gene flow) with more genetic data could provide further information to understand the process of evolution of M- and S-type variations in *C. falcatum* subsp. *littorare*.

## REFERENCE

1. Manton I, others. Problems of cytology and evolution in the Pteridophyta. Probl Cytol Evol Pteridophyta. Cambridge University Press; 1950;
2. Matsumoto S. Chromosome numbers and the reproductive modes of Japanese *Cyrtomium*. Ann Tsukuba Bot Gard. 1985;3: 1–7.
3. Lu J-M, Cheng X, Wu D, Li D-Z. Chromosome study of the fern genus *Cyrtomium* (Dryopteridaceae). Bot J Linn Soc. 2006;150: 221–8.
4. Matsumoto S. Species ecological study on reproductive systems and speciation of *Cyrtomium falcatum* complex (Dryopteridaceae) in Japanese archipelago. Ann Tsukuba Bot Gard. 2003;22: 1–141.
5. Klekowski EJ. Reproductive biology of the Pteridophyta. II. Theoretical considerations. Bot J Linn Soc. 1969;62: 347–59.
6. Masuyama S. The sequence of the gametangium formation in homosporans fern gametophytes: 1. Patterns and their possible effect on the fertilization, with special reference to the gametophytes of *Athyrium*. Sci Rep Tokyo Kyoiku Daigaku, B. 1975;16: 47–69.
7. Barrett SCH, Husband BC. Variation in Outcrossing Rates in *Eichhornia*

- paniculata: The Role of Demographic and Reproductive Factors\*. *Plant Species Biol.* 1990;5: 41–55.
8. Barrett SCH, Ness RW, Vallejo-Marín M. Evolutionary pathways to self-fertilization in a tristylous plant species. *New Phytol.* 2009;183: 546–56.
  9. Ness RW, Siol M, Barrett SCH. Genomic consequences of transitions from cross- to self-fertilization on the efficacy of selection in three independently derived selfing plants. *BMC Genomics. BioMed Central*; 2012;13: 611.
  10. Schoen DJ, Morgan MT, Bataillon T. How Does Self-Pollination Evolve? Inferences from Floral Ecology and Molecular Genetic Variation. *Philos Trans R Soc B Biol Sci.* 1996;351: 1281–90.
  11. Lande R, Schemske DW. the Evolution of Self-Fertilization and Inbreeding Depression in Plants .1. Genetic Models. *Evolution.* 1985;39: 24–40.
  12. Busch JW, Joly S, Schoen DJ. Demographic signatures accompanying the evolution of selfing in *Leavenworthia alabamica*. *Mol Biol Evol.* 2011;28: 1717–29.
  13. Hazzouri KM, Escobar JS, Ness RW, Killian Newman L, Randle AM, Kalisz S, et al. Comparative population genomics in *Collinsia* sister species reveals evidence for reduced effective population size, relaxed selection, and evolution of

- biased gene conversion with an ongoing mating system shift. *Evolution*. 2013;67: 1263–78.
14. Bertorelle G, Benazzo A, Mona S. ABC as a flexible framework to estimate demography over space and time: Some cons, many pros. *Mol Ecol*. 2010;19: 2609–25.
  15. Shepherd LD, McLay TGB. Two micro-scale protocols for the isolation of DNA from polysaccharide-rich plant tissue. *J Plant Res*. 2011;124: 311–4.
  16. Lian CL, Wadud MA, Geng Q, Shimatani K, Hogetsu T. An improved technique for isolating codominant microsatellite markers . *J plant Resour*. 2006;119: 415–7.
  17. Takayama K, Patricio López S, König C, Kohl G, Novak J, Stuessy TF. A simple and cost-effective approach for microsatellite isolation in non-model plant species using small-scale 454 pyrosequencing. *Taxon*. 2011;60: 1442–9.
  18. Lian C, Zhou Z, Hogetsu T. A Simple Method for Developing Microsatellite Markers using Amplified Fragments of Inter-simple Sequence Repeat (ISSR). *J Plant Res*. 2001;114: 381–5.
  19. Rozen S, Skaletsky H. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol*. 2000;132: 365–86.

20. Brownstein MJ, Carpten JD, Smith JR. Modulation of non-templated nucleotide addition by Taq DNA polymerase: primer modifications that facilitate genotyping. *Biotechniques*. 1996;20: 1004–6.
21. Megléc E, Costedoat C, Dubut V, Gilles A, Malausa T, Pech N, et al. QDD: A user-friendly program to select microsatellite markers and design primers from large sequencing projects. *Bioinformatics*. 2009;26: 403–404.
22. Schuelke M. An economic method for the fluorescent labeling of PCR fragments. *Nat Biotechnol*. 2000;18: 233–4.
23. Nei M. *Molecular evolutionary genetics*. Columbia university press; 1987.
24. Mousadik A, Petit RJ. High level of genetic differentiation for allelic richness among populations of the argan tree [*Argania spinosa* (L.) Skeels] endemic to Morocco. *Theor Appl Genet*. 92: 832–9.
25. WRIGHT S. *THE GENETICAL STRUCTURE OF POPULATIONS*. Ann Eugen. Blackwell Publishing Ltd; 1949;15: 323–54.
26. Goudet J. FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9. 3). 2001.
27. Chybicki IJ, Burczyk J. Simultaneous estimation of null alleles and inbreeding coefficients. *J Hered*. 2009;100: 106–13.

28. Hedrick PW. Genetic Load and the Mating System in Homosporous Ferns. *Evolution*. 1987;41: 1282–9.
29. David E. McCauley Dean P. Whittier LMR. Inbreeding and the Rate of Self-Fertilization in a Grape Fern, *Botrychium dissectum*. *Am J Bot. Botanical Society of America*; 1985;72: 1978–81.
30. Soltis PS, Soltis DE. Evolution of Inbreeding and Outcrossing in Ferns and Fern-Allies. *Plant Species Biol*. 1990;5: 1–11.
31. David C. Queller KFG. Estimating Relatedness Using Genetic Markers. *Evolution*. 1989;43: 258–75.
32. B. S. Weir CCC. Estimating F-Statistics for the Analysis of Population Structure. *Evolution*. 1984;38: 1358–70.
33. Meirmans PG, Hedrick PW. Assessing population structure: FST and related measures. *Mol Ecol Resour*. 2011;11: 5–18.
34. PE PRS. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research?an update. *Bioinformatics*. 2012;28: 2537–2539.
35. Wright S. Isolation by Distance. *Genetics*. 1943;28: 114–38.
36. Rousset F. Genetic Differentiation and Estimation of Gene Flow from F-Statistics Under Isolation by Distance. *Genetics*. 1997;145: 1219–28.

37. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics*. 2000;155: 945–59.
38. Hubisz MJ, Falush D, Stephens M, Pritchard JK. Inferring weak population structure with the assistance of sample group information. *Mol Ecol Resour*. 2009;9: 1322–32.
39. Jakobsson M, Rosenberg NA. CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*. 2007;23: 1801–6.
40. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol Ecol*. 2005;14: 2611–20.
41. Earl DA, vonHoldt BM. STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour*. 2012;4: 359–61.
42. Bruno WJ, Socci ND, Halpern L. Weighted neighbor joining: a likelihood-based approach to distance-based phylogeny reconstruction. *Mol Biol Evol*. 2000;17: 189–97.
43. Cornuet JM, Santos F, Beaumont M, Robert CP, Marin JM, Balding DJ, et al.

- Inferring population history with DIY ABC: A user-friendly approach to approximate Bayesian computation. *Bioinformatics*. 2008;24: 2713–9.
44. Cornuet JM, Pudlo P, Veyssier J, Dehne-Garcia A, Gautier M, Leblois R, et al. DIYABC v2.0: A software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. *Bioinformatics*. 2014;30: 1187–9.
  45. Estoup A, Jarne P, Cornuet JM. Homoplasmy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Mol Ecol*. 2002;11: 1591–604.
  46. Garza JC, Williamson EG. Detection of reduction in population size using data from microsatellite loci. *Mol Ecol*. 2002;10: 305–18.
  47. Excoffier L, Estoup A, Cornuet JM. Bayesian analysis of an admixture model with mutations and arbitrarily linked markers. *Genetics*. 2005;169: 1727–38.
  48. Sakaguchi S, Bowman DMJS, Prior LD, Crisp MD, Linde CC, Tsumura Y, et al. Climate, not Aboriginal landscape burning, controlled the historical demography and distribution of fire-sensitive conifer populations across Australia. *Proc Biol Sci*. 2013;280: 20132182.
  49. Wubs ERJ, de Groot GA, During HJ, Vogel JC, Grundmann M, Bremer P, et al.

- Mixed mating system in the fern *Asplenium scolopendrium*: implications for colonization potential. *Ann Bot.* 2010;106: 583–90.
50. Soltis DE, Soltis PS. Breeding System of the Fern *Dryopteris expansa*: Evidence for Mixed Mating. *Am J Bot. Botanical Society of America*; 1987;74: 504–9.
51. Soltis PS, Soltis DE. Estimated Rates of Intragametophytic Selfing in Lycopods. *Am J Bot. Botanical Society of America*; 1988;75: 248–56.
52. Ranker T. Genetic diversity, mating systems, and interpopulation gene flow in neotropical *Hemionitis palmata* L. (Adiantaceae). *Heredity.* 1992;69: 175–83.
53. Ranker T, Gemmill CE, Trapp PG. Microevolutionary patterns and processes of the Native Hawaiian colonizing fern *Odontosoria chinensis* (Lindsaeaceae). *Evolution.* 2000;54: 828–39.
54. De Groot G, During HJ, Ansell SW, Schneider H, Bremer P, Wubs ERJ, et al. Diverse spore rains and limited local exchange shape fern genetic diversity in a recently created habitat colonized by long-distance dispersal. *Ann Bot.* 2012;109: 965–78.
55. Chung MY, Moon MO, López-Pujol J, Maki M, Yamashiro T, Yukawa T, et al. Was Jeju Island a glacial refugium for East Asian warm-temperate plants? insights from the homosporous fern *Selliguea hastata* (Polypodiaceae). *Am J Bot.*

- 2013;100: 2240–9.
56. Chung MY, López-Pujol J, Chung JM, Kim K-J, Chung MG. Low Within Population Genetic Variation and High Among Population Differentiation in *Cyrtomium falcatum* (Lf) C. Presl (Dryopteridaceae) in Southern Korea: Inference of Population-Establishment History. *Am Fern J.* 2012;102: 256–72.
57. Hamrick JL, Godt MJW, Brown AHD, Clegg MT, Kahler AL, Weir BS, et al. Allozyme diversity in plant species. *Plant Popul Genet breeding, Genet Resour.* 1990; 43–63.
58. Pollak E. On the theory of partially inbreeding finite populations. I. Partial selfing. *Genetics.* 1987;117: 353–60.
59. Shimizu KK, Tsuchimatsu T. Evolution of Selfing: Recurrent Patterns in Molecular Adaptation. *Annu Rev Ecol Evol Syst.* 2015;46: 593-622.
60. Nordborg M. Linkage Disequilibrium, Gene Trees and Selfing: An Ancestral Recombination Graph With Partial Self-Fertilization. *Genetics.* 2000;154: 923–9.
61. Baker HG. Self-Compatibility and Establishment After “Long-Distance” Dispersal. *Evolution.* 1955;9: 347–9.
62. Watano Y. Genetic life history of the homosporous fern *Osmunda lancea*. *Mod Asp species Univ Tokyo Press Tokyo.* 1986; 211–9.

63. Yuzawa Y. Assessment of damages on coastal and seashore plants in Fukushima Prefecture caused by Japan Tsunami following the Great East Japan Earthquake on 2011. *J Phytogeogr Taxon.* 2012;61: 1–14.
64. Jiménez A, Holderegger R, Csencsics D, Quintanilla LG. Microsatellites reveal substantial among-population genetic differentiation and strong inbreeding in the relict fern *Dryopteris aemula*. *Ann Bot.* 2010;106: 149–55.
65. Iwaizumi MG, Tsuda Y, Ohtani M, Tsumura Y, Takahashi M. Recent distribution changes affect geographic clines in genetic diversity and structure of *Pinus densiflora* natural populations in Japan. *For Ecol Manage.* 2013;304: 407–16.
66. Tsuda Y, Ide Y. Wide-range analysis of genetic structure of *Betula maximowicziana*, a long-lived pioneer tree species and noble hardwood in the cool temperate zone of Japan. *Mol Ecol.* 2005;14: 3929–41.
67. Tsuda Y, Sawada H, Ohsawa T, Nakao K, Nishikawa H, Ide Y. Landscape genetic structure of *Betula maximowicziana* in the Chichibu mountain range, central Japan. *Tree Genet Genomes.* 2010;6: 377–87.
68. Bystriakova N, Ansell SW, Russell SJ, Grundmann M, Vogel JC, Schneider H. Present, past and future of the European rock fern *Asplenium fontanum*:

- combining distribution modelling and population genetics to study the effect of climate change on geographic range and genetic diversity. *Ann Bot.* 2014;113: 453–65.
69. Tsukada M. Vegetation and climate during the last glacial maximum in Japan. *Quat Res.* 1983;19: 212–35.
70. Tsuda Y, Ide Y. Chloroplast DNA phylogeography of *Betula maximowicziana*, a long-lived pioneer tree species and noble hardwood in Japan. *J Plant Res.* 2010;123: 343–53.
71. Hu LJ, Uchiyama K, Shen HL, Saito Y, Tsuda Y, Ide Y. Nuclear DNA microsatellites reveal genetic variation but a lack of phylogeographical structure in an endangered species, *Fraxinus mandshurica*, across north-east China. *Ann Bot.* 2008;102: 195–205.
72. Tsuda Y, Nakao K, Ide Y, Tsumura Y. The population demography of *Betula maximowicziana*, a cool-temperate tree species in Japan, in relation to the last glacial period: its admixture-like genetic structure is the result of simple population splitting not admixing. *Mol Ecol.* 2015;24: 1403–18.
73. Sato T, Sakai A. Distributional patternizations and characteristics of Japanese Pteridophyta. *Low Temp Sci Ser B Biol Sci.* 1978;

74. Igarashi Y, et al. Vegetation history of Kenbuchi Basin and Furano Basin in Hokkaido, north Japan, since 32 000 yrs BP. *Quat Res.* 1993;32: 89–105.
75. Kitamura K, Matsui T, Kobayashi M, Saitou H, Namikawa K, Tsuda Y. Decline in gene diversity and strong genetic drift in the northward-expanding marginal populations of *Fagus crenata*. *Tree Genet Genomes.* 2015;11.

Table 1.1. Primer sequences, repeat motifs, and accession numbers of source sequences for eight microsatellite markers developed in this study.

Locus	Primer sequence (5'–3')	Repeat motif	Accession number
CFL-079*	F: pig1-ACGAAGAAGGACGAGTAGC	(AC) <sub>6</sub> (AG) <sub>10</sub>	LC055975
CFL-079*	R: ACACACACACAGAGAGAGAGAG	(AC) <sub>6</sub> (AG) <sub>10</sub>	LC055975
CFL-C32*	F: CTCATGGGACTTTTTGTGTC	(AC) <sub>6</sub> (AG) <sub>10</sub>	LC055976
CFL-C32*	R: ACACACACACAGAGAGAGAGAG	(AC) <sub>6</sub> (AG) <sub>10</sub>	LC055976
CFL-Z03#	F: u19-AATGGAAGAGGGCACGAGTA	(GA) <sub>8</sub>	LC055977
CFL-Z03#	R: pig2-GCATGTCCAAAGGAGTGACTT	(GA) <sub>8</sub>	LC055977
CFL-B02#	F: u19-GCTTGCTTGACAGAGACACG	(GA) <sub>14</sub> , A <sub>7</sub>	LC055978
CFL-B02#	R: pig2-TATGAACGGATAGTGCCACG	(GA) <sub>14</sub> , A <sub>7</sub>	LC055978
CFL-B12#	F: u19-CCGTTGAAGGTGGGAAGTAA	(TC) <sub>12</sub>	LC055979
CFL-B12#	R: pig2-AGCCTCCATGCCTCCTTTAT	(TC) <sub>12</sub>	LC055979
CFL-B13#	F: u19-TCGGCTCTACCTCCTCTCAA	(AC) <sub>11</sub>	LC055980
CFL-B13#	R: pig2-ATGAGTGCATATGGGCAACA	(AC) <sub>11</sub>	LC055980
CFL-B16#	F: u19-GTAAGTGGGCACTTTCCTG	(GT) <sub>11</sub>	LC055981
CFL-B16#	R: pig2-GCGCTAAGGTTGTTTCGTCTC	(GT) <sub>11</sub>	LC055981
CFL-B17#	F: u19-GACGAGGGCGTAAATGAGAA	(TC) <sub>5</sub> , (AC) <sub>11</sub> , A <sub>4</sub>	LC055982
CFL-B17#	R: pig2-GCCATAACGTCAAGGCAAGT	(TC) <sub>5</sub> , (AC) <sub>11</sub> , A <sub>4</sub>	LC055982

\*, markers developed by the method of Lian *et al.* (2006); #, markers developed by next generation sequencing.

u19, 5'-GGTTTTCCCAGTCACGACG-3'; pig1, 5'-GTTTCTT-3'; pig2, 5'-GTTT-3'.

Table 1.2. Genetic diversity indices and inbreeding coefficient values for seven populations of *Cyrtomium falcatum* subsp. *littorale*.

	M_type populations					S_type populations			
	ESAN1	ESAN2	SAND	KANT	Mean	IZU1	IZU2	SADO	Mean
$N_A$	2.13	2.125	2.5	2.25	2.251	3.75	2.625	3	3.167
$A_R$	1.7	1.847	2.38	2.074	2	2.9	2.39	2.863	2.718
$h$	0.076	0.158	0.227	0.217	0.17	0.378	0.326	0.436	0.38
$F_{IS}^*$	<b>0.501</b>	<b>0.671</b>	<b>0.794</b>	<b>0.56</b>	0.632	<b>0.34</b>	<b>0.22</b>	-0.1	0.153
$N_1$	169	210	299	328	215.5	846	465	715	675.3

Bold type indicates significant  $F_{IS}$  values ( $p < 0.00089$ , simple Bonferroni correction for 5% level); \*, multi locus estimate.

$N_A$  is mean number of alleles.  $A_R$ , Allelic richness, and  $h$ , genediversity, are indexes of genetic diversity. The  $F_{IS}$  is Wright's fixation index, that index reflects selfing rate.  $N_1$  is current population size estimated by DIYABC.

Table 1.3. Group comparison of genetic variation between M type and S type evaluated for mean of allelic richness,

observed heterozygosity ( $H_o$ ), gene diversity ( $h$ ), fixation index ( $F_{is}$ ), relatedness and  $F_{ST}$ .

*P*-values are based on the permutation test implemented in FSTAT.

	<i>allelic richness</i>	<i>h</i>	$F_{is}$	relatedness	$F_{ST}$
M-type	1.999	0.152	0.626	0.708	0.663
S-type	2.718	0.367	0.208	0.490	0.367
<i>P</i> -value (M>S)	0.989	1.000	0.029*	0.064	0.067
<i>P</i> -value (S>M)	0.016*	0.009*	0.981	0.963	0.967

Table 1.4. Inbreeding coefficient ( $F_{is}$ ) values estimated by INEST2 for seven populations of *Cyrtomium falcatum* subsp. *littorale*.

	M_type population					S_type population			Mean
	ESA N1 <i>n</i> =41	ESA N2 <i>n</i> =36	SAN D <i>n</i> =16	KAN T <i>n</i> =28	Mea n <i>n</i>	IZU1 <i>n</i> =42	IZU2 <i>n</i> =35	SAD O <i>n</i> =17	
$F_{is}$ mean	0.484	0.634	0.772	0.478	0.592	0.274	0.225	0.019 7	0.173
Low (95%)	0.205	0.435	0.567	0.211	0.355	0.140	0.210	0.000	0.117
High (95%)	0.720	0.815	0.942	0.699	0.794	0.408	0.372	0.067 2	0.282

$F_{is}$  (Fixation index) estimated by INEST2. INEST2 estimates FIS with correcting effect of null alleles.

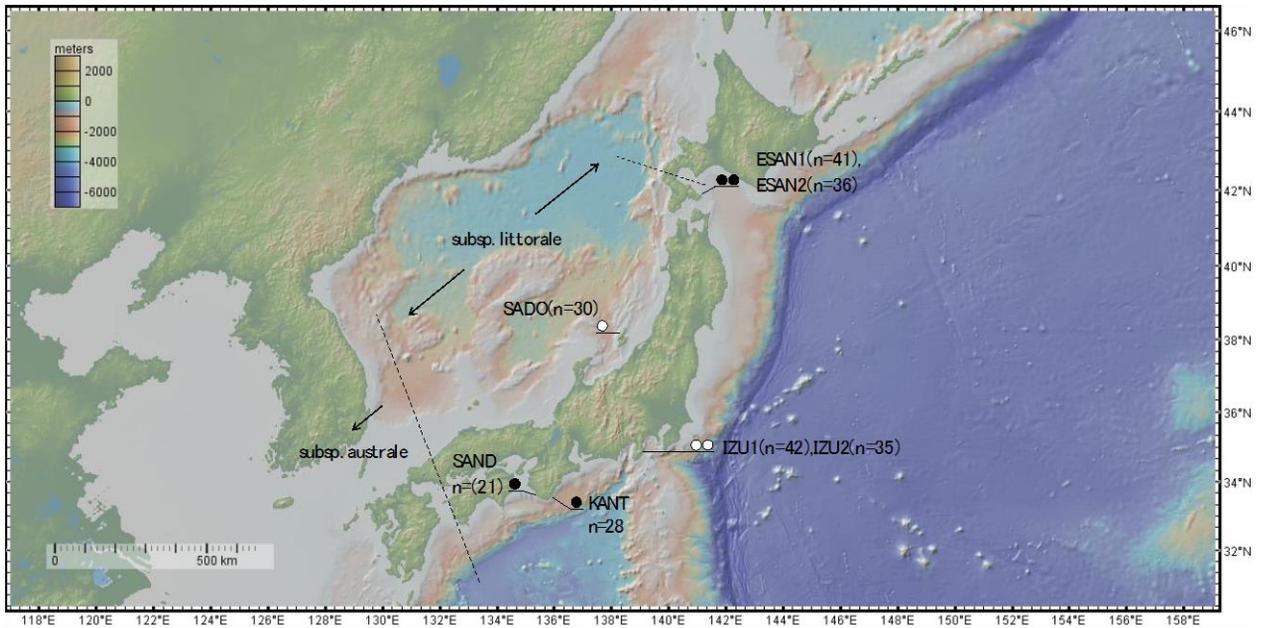


Fig. 1.1 The northern and southern limits of distributional range of *C. falcatum* subsp. *littorale* based on Matsumoto (2003), and sampling locations of 7 populations examined in this study. Black dots are M type and white dots are S type populations of *Cyrtomium falcatum* subsp. *littorale* sampled in this study. Dashed lines are northern and southern limit of *C. falcatum* subsp. *littorale*. And southern limit is faced with distribution of *C. falcatum* subsp. *australe*.

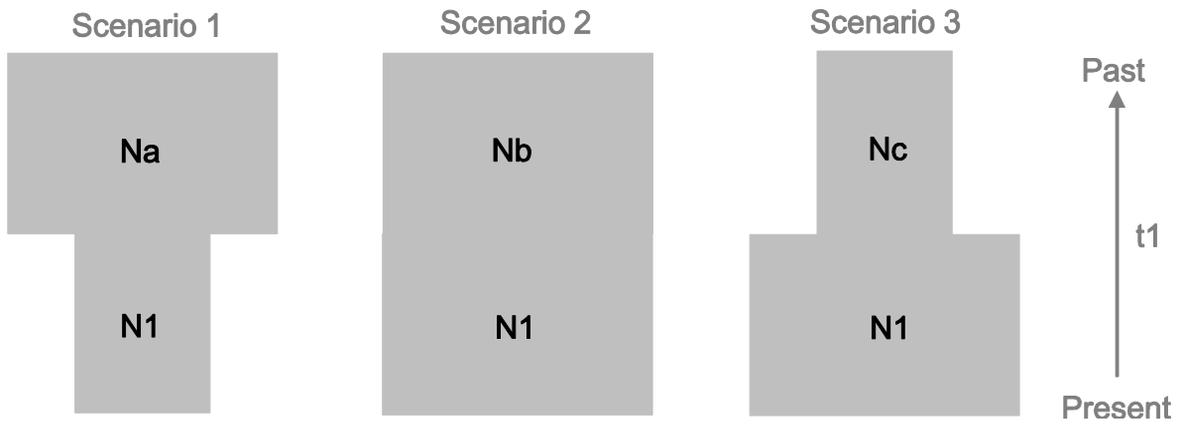


Fig. 1.2 Demographic scenarios for each population of *C. falcatum* subsp *littorale* used in DIY ABC.

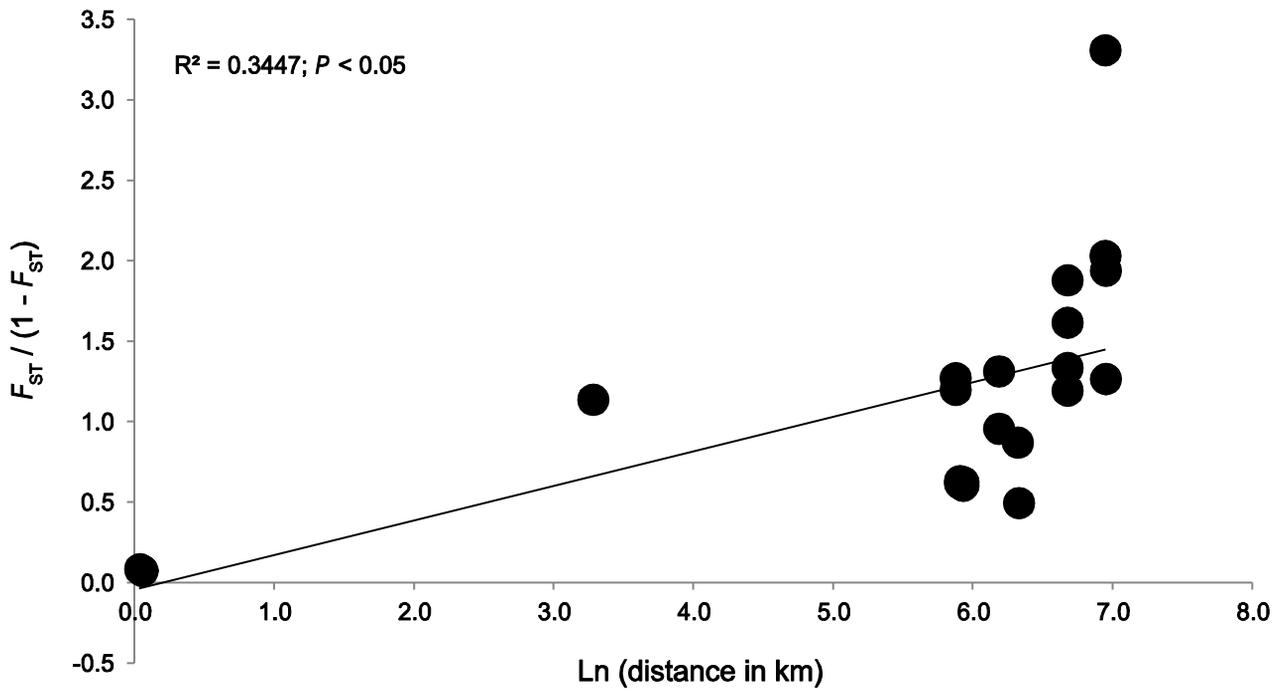
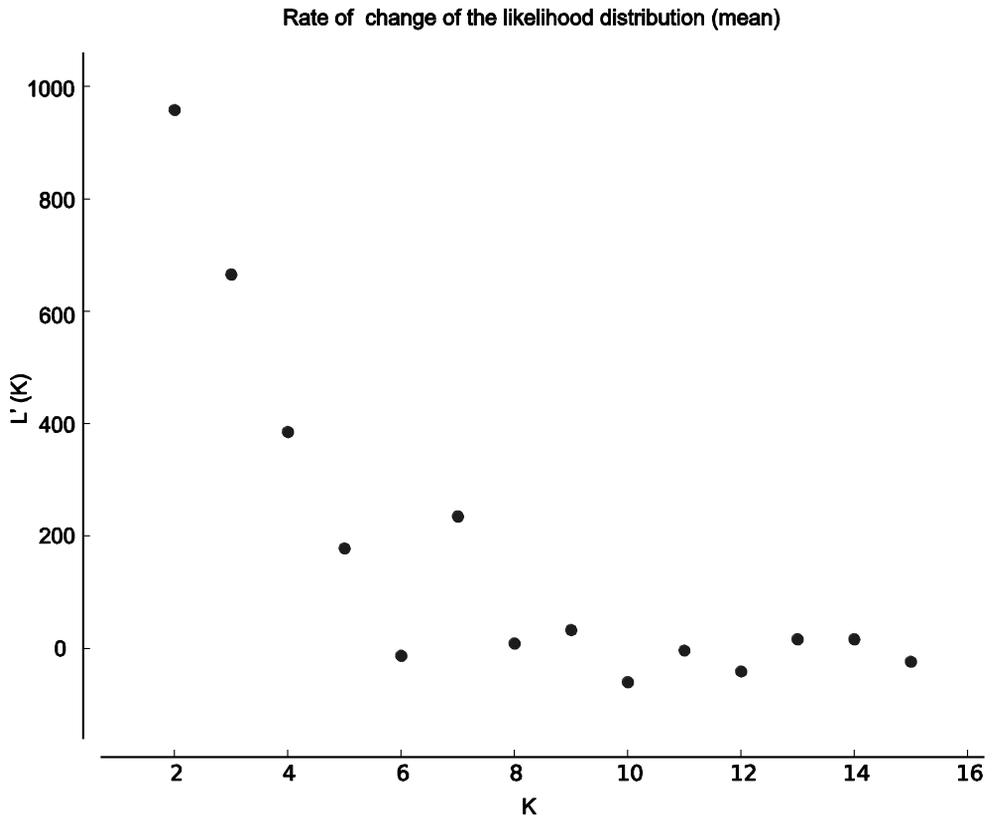


Fig. 1.3 Isolation by distance for the 7 populations of *C. falcatum* subsp *littorale*  
 The relationship between the matrix of pairwise differentiation described as  $F_{ST} / (1 - F_{ST})$  and the matrix of the natural logarithm of geographic distance (in meters) in the 7 populations

A



B

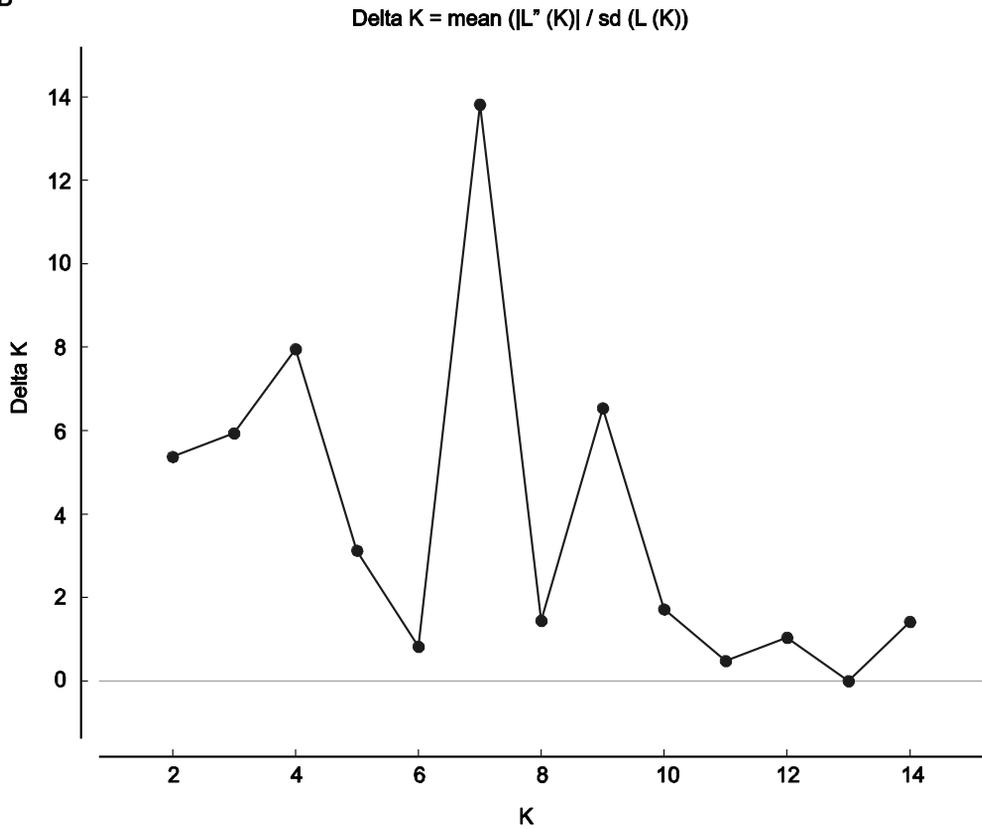
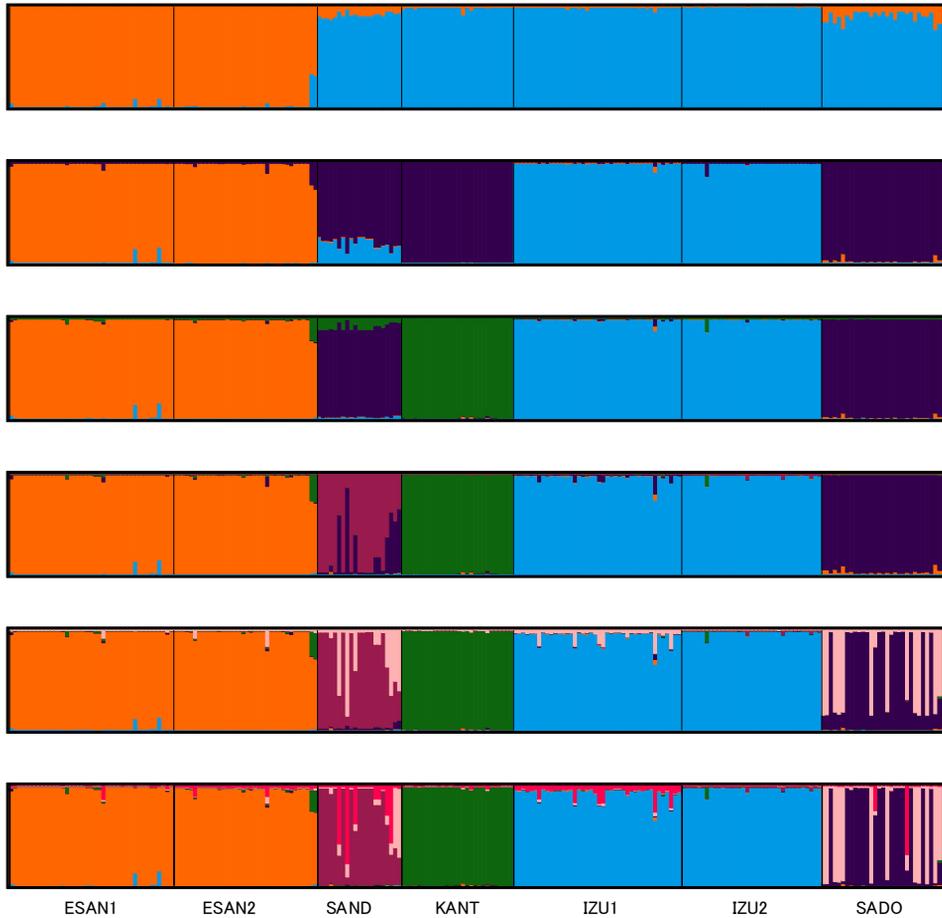


Fig. 1.4 The values of posterior probability of the data ( $\ln P(D)$ ) from 10 runs for each value of  $K$  (1 – 15; A) and  $\Delta K$  (right B).

A



B

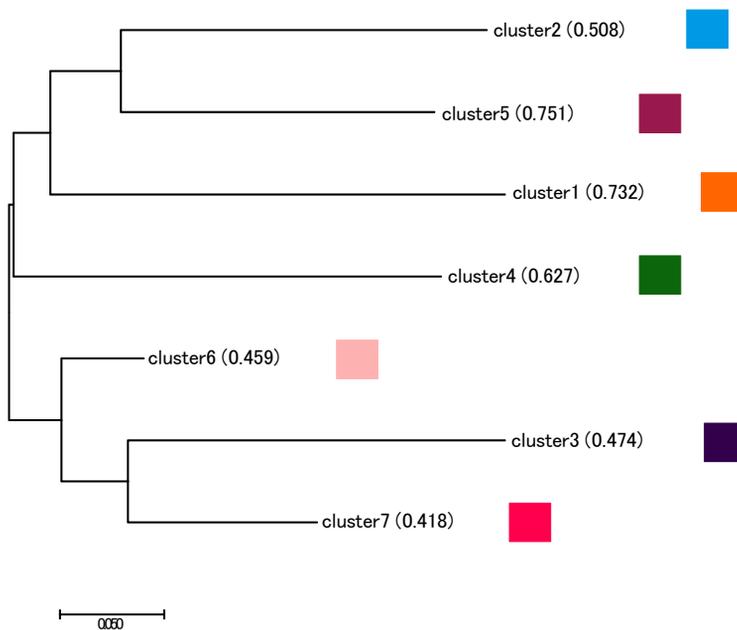


Fig. 1.5 Results of STRUCTURE analysis and Neighbor-joining (NJ) tree of each clusters in  $K = 7$

The proportion of the membership coefficient of 233 individuals in the 7 populations for each of the inferred clusters for  $K = 2-7$  defined using Bayesian clustering in STRUCTURE analysis (A). The neighbor-joining tree of the seven clusters for  $K = 7$  (B). Values indicated subsequent to each cluster number are  $F_{ST}$  between each cluster and the common ancestral population.

## **CHAPTER 2: Divergence of *Cyrtomium falcatum* subsp. *littorale* and the ancestral subsp. *australe*: phylogenetic inferences from RAD-seq data**

### **INTRODUCTION**

In the previous chapter, mating system divergence among *Cyrtomium falcatum* subsp. *littorale* populations were reported. The M-type populations showed significantly larger  $F_{IS}$ , smaller gene diversity and smaller effective population size than the S-type populations. Previous chapter also showed mixed mating in *C. falcatum* subsp. *littorale* except for SADO population. Matsumoto (2003)[1] suggested that *C. falcatum* subsp. *littorale* could have been derived from *C. falcatum* subsp. *australe* because the former has more specialized characters, such as the mixed type gametangium formation (adapted for selfing) and lithophytic habit (growing on rock), than the latter. However, the phylogenetic relationship between populations of the two sexual subspecies has not yet been well examined in detail. In order to understand mating system evolution in *C. falcatum*, therefore, it is important to reconstruct reliable history of divergence among populations of the two subspecies. As shown in chapter 1, the resolution power of the microsatellite markers we developed was not high enough to evaluate genetic relatedness among the populations of subsp. *littorale* in detail (Fig. 1.5, page 46). The recent development of next generation sequencing (NGS) technologies has enabled us to generate genome-wide single nucleotide polymorphisms (SNPs) rapidly. Restriction site associated sequencing (RAD-seq) is one of the NGS methods applicable to non-model organisms. In the present study, a huge nucleotide sequence dataset generated by single-end double digest RAD-seq method was used for phylogenetic and demographic analyses of sexual diploid *C. falcatum*.

The purposes of this chapter are 1) to show phylogenetic relationship among

populations of the two sexual diploid subspecies *littorale* and *austral* by using SNPs, and 2) to know how evolutionary transition(s) in mating system has(have) occurred in sexual diploid populations of *C. falcatum*.

## MATERIALS AND METHODS

### *Preparation of samples*

A total of 23 samples of *Cyrtomium falcatum* subsp. *australe* were newly collected from four populations. Abbreviations of the four populations and their localities were as follows: **TUNO** (34° 21' 5.8" N, 130° 50' 19.9" E) in Tsuno-shima, Yamaguchi Pref.; **NOMO** (32°35' 39.2" N, 129°45' 44.0" E) in Nomo-saki, Nagasaki Pref.; **NAKA** (32°55' 16.9" N, 129°00' 22.0" E) in Nakadori, Nagasaki Pref.; and **KASA** (31°25' 34.1" N, 130°08' 44.5" E) in Kasasa-cho, Kagoshima Pref. (Fig. 2.1). As for *C. falcatum* subsp. *littorale*, I employed a subset (26) of the samples used in chapter 1: five from ESAN1, five from ESAN2, eight from SAND, and eight from SADO. Although I used twelve samples from IZU1, DNAs of these samples were extracted from the fresh samples collected at the same location. This is because quality of DNAs extracted from the dried leaf samples of IZU1 were not enough to be used for RAD-seq. Additionally, I used 23 individuals (14 *C. f. littorale* and nine *C. f. australe*) cultivated in Tsukuba Botanical Gardens, which was collected and used in Matsumoto (2003)[1]: one from ESAN, three from IZU (IZU1 or IZU2), two from SAND and six subsp. *australe* individuals from Ogasawara Islands, Tokyo (**OGAS**). The rest samples were used to preliminary samples. In addition to the sporophyte samples, twelve haploid gametophyte samples were also used. The gametophytes are progeny of the hybrid between the two subspecies (A1-55 x A2-2), which was made through artificial crossing by Matsumoto (2003). The sequence data from these haploid samples were also used to remove contigs containing paralogous sequences. Finally a total of 96 samples were subjected to RAD-seq (Table 2.1).

Concentration of each DNA samples was measured using Qubit and its assay kit (Life Technologies Japan, Tokyo) and adjusted to 20ng/μL with TE buffer. The samples were sent to Center for Ecological Research, Kyoto University and were processed there.

### *RAD-sequencing*

RAD sequence was conducted in Center for Ecological Research, Kyoto University as collaborative research. The method employed was a variety of RAD sequencing, double-digest RAD-seq (ddRAD-seq), and two restriction enzymes, *EcoRI* and *BglII*, were selected. The library constructed was sequenced on Hiseq (Illumina), at one lane in single end 50bp reads mode.

### *SNP discovery and filtering of the data*

The fastq file, in which sequences were sorted to individuals according to the barcode sequences made in Center for Ecological Research, were processed using pyRAD [2] to construct contigs. The low quality reads (quality value Q less than 33) were discarded. At the step of within-sample clustering, minimum coverage (number of reads per sample), maximum number of heterozygous sites and the allowed number of alleles were set to 5, 5 and 2, respectively. After across-samples clustering, the loci that have data of at least 76 samples (76/96) were retained. Variant Call Format (VCF) file were used to following analysis that include only SNPs generated by pyRAD. As a first step of data filtering, I removed loci that contain < 20% missing data since missing data can bias data analysis. Next, I removed loci that have < 5 depth or only minor alleles < 5%. This is because Roesti et al. (2012)[3] suggested that minor alleles are likely to be uninformative and can bias genome scan (e.g. outlier loci detection) which is important to evaluate both selection and neutral genetic structure [4]. The minor allele frequency threshold with 5% is widely employed in recent NGS-based studies [5–7]. To detect outlier loci, I used three approaches of genome scan. Since these filtering methods are based on allele frequency data of populations, I used nine populations that contained more than four individuals for following analyses. First, I used BayeScan 2.1 [8–10]. This software identifies candidate loci under selection using differences in allele

frequencies between populations. The run was conducted on following parameters; burn in: 50000; thinning interval: 10; sample size: 5000; resulting total number of iterations: 100000; number of pilot runs: 20; length of each pilot run: 5000. Second, I used Arlequin 3.5.2.2 [11] and this is basically based on the  $F_{ST}$  outlier loci analysis [12] under the assumption of the infinite allele model. The program was run on hierarchical island model and following parameters; number of simulations: 20000; number of demes to simulate: 10; number of groups to simulate: 10; minimum expected heterozygosity: 0; maximum expected heterozygosity: 1, all populations were treated as same hierarchy and assumed no large scale structure. Finally, I used TESS3 [13]. This software can detect outlier loci considering spatial genetic structure. The program was run for nine populations using dataset including location information (longitude and latitude information) of individuals. As discussed in Luikart et al. (2003)[4] and Antao et al. (2008)[14], adaptive outlier loci would cause bias in the population genetics analysis which is intended to evaluate neutral genetic structure and assumes neutrality. Thus, to be conservative about marker selection, I used loci that were assumed as neutral in all three software in the following data analysis.

#### *Basic population genetic analysis*

Average expected heterozygosity ( $H_e$ ), allele frequencies,  $F_{IS}$  and  $F_{ST}$  values at each SNP locus were calculated by using adegenet 1.3 [15].

#### *Cluster analysis*

To recognize population structure, I conducted the cluster analysis using sNMF 1.2 [16]. This software estimates ancestry coefficient of individuals, which was used to recognize intra-population structure or genetic clusters in large scale. The program was run under cluster number  $K = 1$  to 20 with the data masked randomly in 5%, and the optimal cluster number was chosen by cross entropy method using sNMF.

### *Phylogenetic analyses*

All neutral SNPs were concatenated and the resultant data matrix was edited as a file of Phylip format. Heterozygous sites were indicated by IUPAC ambiguity nucleotide codes. The data of gametophyte samples were removed for Phylogenetic analyses.

Phylogenetic tree were inferred using maximum likelihood (ML) method implemented in RAxML 8.0.0 [17]. GTR + gamma model of sequence evolution was employed because the model was usually used in multi-locus sequence data generated by NGS. Support values for branches were estimated through bootstrap analysis of 1000 replicates.

### *Population admixture analysis*

Admixture of clusters were observed in cluster analysis, therefore I verified source population of these admixtures. I used Treemix 1.12 [18] to reconstruct the patterns of population splits and mixtures in multiple populations. Treemix infers the places where mixtures between populations happened in population tree under the assumed number of admixture events. I assumed 1-10 admixture events and used the –global option which is the option to performe a round of global rearrangements of the graph after initial fitting. The best tree was defined by likelihood and decreasing of standard error [19].

## RESULTS

### *De novo assembly of NGS reads*

A total of 181 million reads of 50 bp from 96 samples was obtained from one lane of HiSeq. Over Q30 reads (base call accuracy 99.9%) in a sample was 96.5% on average. A total of 9,337 SNPs were called. The number of contigs or putative loci (hereafter, called simply as loci) meeting the qualifications and containing at least one SNP site were 3,743.

After removing minor alleles, 2,654 loci remained. TESS3 estimated 343 loci, Arlequin 3.5.2.2 estimated 521 loci and BayeScan 2.1 estimated 46 loci as non-neutral loci. 1777 loci were used to following analysis as a result of these filtering (Fig. 2.S1).

### *Basic population genetic analysis*

The expected heterozygosity and FIS values in nine populations examined were given in Table 2.2. The expected heterozygosity values was lower in three subsp. *littorale* populations (ESAN, IZU and SAND) than SADO of subsp. *littorale* and five subsp. *australe* populations (KASA, NAKA, NOMO, TUNO, OGAS). The  $F_{IS}$  values of three subsp. *littorale* populations (ESAN, IZU and SAND) and OGAS were larger than zero (0.180 to 0.472), while nearly zero or negative in the other populations (-0.895 to 0.057). Pairwise  $F_{ST}$  values were shown in Table 2.3. Pairwise  $F_{ST}$  values between populations of subsp. *littorale* (0.370 - 0.655) were much higher than those between populations of subsp. *littorale* except OGAS (-0.095 - 0.233). Interestingly,  $F_{ST}$  values between SADO and each of three subsp. *littorale* populations (0.370 - 0.655) were higher than those between SADO and each of subsp. *australe* populations (0.042 - 0.261).

### *Phylogenetic relationship among populations*

The ML unrooted tree indicated three distinct groups in sexual diploid

populations of *Cyrtomium falcatum* (Fig. 2.2). Most samples of subsp. *littorale* except those of SADO (Niigata Pref.) formed a group. Five samples from Ogasawara Islands (subsp. *australe*) also formed a well-supported group. All samples of SADO and two of SAND (subsp. *littorale*), and one of Ogasawara Islands (subsp. *australe*) were grouped with those collected from Yamaguchi Pref. and the Kyushu Island (subsp. *australe*). Hereafter, the three groups are referred as LITTORALE, OGASAWARA and KYUSHU, respectively.

#### *Cluster analysis*

The cross entropy steadily decrease from  $K=1$  to 7 and the smallest value was obtained at  $K = 7$  (Fig. 2.S2), suggesting that  $K = 7$  is the optimal cluster number. In  $K = 2$ , subsp. *littorale* and *australe* were not divided clearly. IZU and ESAN were cluster 1, and NOMO and TUNO belonged cluster 2. The other populations were composed of the individuals of the two clusters or admixed individuals. In  $K = 3$ , NOMO and TUNO were assigned to cluster 3. In  $K = 4$ , OGAS was distinguished from the other subsp. *australe* populations. In  $K = 5$ , ESAN, IZU and SAND were divided to its own cluster. In  $K = 6$ , ESAN and IZU were grouped in the same cluster. In  $K = 7$ , ESAN, IZU and SAND were assigned to its own cluster again. SADO was composed of individuals belonging to the cluster unique to SADO and admixed individuals of IZU and NAKA clusters (Fig. 2.3).

#### Population tree with migration edges

The likelihood of trees steadily increased to assumed migration number  $m = 6$  (Fig. 2.S3). However, the maximum standard error of drift parameter between pairs of populations did not decrease after  $m=3$  (Fig. 2.S4). Therefore,  $m=3$  is considered to be optimum number of migration. Population tree with three migration number  $m=3$  was shown in Fig. 2.4. Trees of  $m = 0$  and 6 were also given in Fig. 2.4. In  $m = 3$ , the 3

migrations edges were identified as follows: 1) LITTORALE to SADO, 2) LITTORALE to KASA, and 3) ancestral population of NAKA to TUNO (Fig. 2.4 B). The three migration edges in  $m=3$  were recovered even in  $m=6$ .

## DISCUSSION

### *Non-monophyly of *Cyrtomium falcatum* subsp. littorale*

The phylogenetic analysis of the two sexual diploid subspecies suggested that individuals of the two subspecies are not reciprocally monophyletic (Fig. 2.2). Individuals of SADO population was separated from other members of subspecies *littorale* (LITTORARE group in this study), and some of them was apparently nested in a group of individuals from Yamaguchi Pref. and Kyushu (KYUSHU group in this study). one of the two S\_type populations as mentioned in chapter 1. The  $F_{IS}$  value of SADO (0.157) was the lowest among the populations examined, and not significantly deviated from zero if admixed individuals shown by STRUCTURE analysis (Fig. 1.5) were excluded (Table 1.2). Furthermore, INEST2 program that calculate  $F_{IS}$  values corrected for null microsatellite alleles suggested that SADO would be outcrossing population (Table 1.3). The low or no selfing rate in SADO would be unique among the populations of subsp. *littorale*. Given that SADO is more closely related to KYUSHU of subsp. *australe*, however, the low or no selfing rate in SADO would be reasonable. Subspecies *littorale* is distinguished from subsp. *australe* by the characters of smaller blades, fewer pairs of pinnae, and grayish indusia without a blackish brown center. The phenotypic character of small and simple leaves may be easily obtained in lithophytic habitat. Additionally, the plants having indusia without a blackish brown center are sometimes observed even in subsp. *australe* (Matsumoto 2003). Possibly, SADO population was derived from KYUSHU lineage of subsp. *australe* independently of the other subsp. *littorale* populations.

### *Genetic structure and admixture in diploid *Cyrtomium falcatum**

In the previous chapter using eight microsatellite markers, the cluster analysis by STRUCTURE (Fig. 1.5) suggested that SAND and SADO contained individuals with odd genetic characters, and I postulated that subsp. *australe* was inadvertently

included in, and admixed with, our samples, especially in the SAND and SADO populations. Cluster analysis by sNMF using 1,777 SNPs loci (Fig. 2.3) also supported this interpretation. Particularly, all of the individuals in SADO were found to be more genetically similar to KASA of subsp. *australe*. Individuals of *C. falcatum* subsp. *littorale* except SADO and some of SAND were assigned to the same cluster from K=2 to 4. Although ESAN, IZU and SAND were assigned to different clusters in K=7 which would be optimum, this would reflect high  $F_{ST}$  values between pairs of subsp. *littorale* (Table 2.2).

As for KYUSHU group, NOMO and TUNO form a cluster thoroughly from K=2 to 7. While, KASA was shown to contain admixed individuals (K=2-5, 7) or heterogeneous individuals (K=6). Even OGAS was composed of admixed individuals in K=2, 3 and 5. The presence of admixed individuals in SADO and most populations of subsp. *australe* suggests that ongoing and historical hybridization events between the two subspecies. For example, one SAND (sand18) individual was nested in the tight cluster of NOMO and TUNO in the phylogenetic tree (Fig. 2.2). The cluster analysis also assign sand18 to the same cluster of NOMO and TUNO thoroughly (Fig. 2.3). This clearly indicates an episode of very recent migration of “typical” subsp. *australe* individual to the Kii Peninsula, the geographic region of subsp. *littorale*. In the phylogenetic tree (Fig. 2.2), four SADO (sado8, 25, 29 and 31), three KASA (kasa3, 4 and 5), one TUNO (tuno8) and one SAND (sand16) individuals stemmed from the middle positions of the edge connecting core KYUSHU (NOMO, TUNO and NAKA) and LITTORALE. The admixture (hybridity) between two subspecies were well shown for these individuals even in the cluster analyses of K=3 (Fig. 2.3).

In Population trees by Treemix (Fig. 2.4), three groups (LITTORALE, OGASAWARA and KYUSHU) were recognized in  $m=0$  and 3. This is consistent with the topology of ML tree (Fig. 2.2). Additionally, SADO formed a group with KYUSHU. This is also concordant with the results of ML tree (Fig. 2.2) and cluster analyses (Fig.

2.3). In  $m=3$  which would be optimal, migration edges from LITTORALE to SADO and from LITTORALE to KASA were inferred. This results is interesting because unidirectional gene flow to subsp. *australe* would be responsible for genetic admixture of the two population shown by other analyses. Additional studies are needed to determine whether the indirection gene flow would be general between the two subspecies.

Individuals of NAKA and OGAS population showed relatively distant each other in phylogenetic tree. NAKA and OGAS population are in small islands isolated from main land. Although the effect of isolation is unclear, this pattern is interesting for future study.

## REFERENCE

1. Matsumoto S. Species ecological study on reproductive systems and speciation of *Cyrtomium falcatum* complex (Dryopteridaceae) in Japanese archipelago. *Ann Tsukuba Bot Gard.* 2003;22: 1–141.
2. Eaton DAR. PyRAD: Assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics.* 2014;30: 1844–9.
3. Roesti M, Salzburger W, Berner D. Uninformative polymorphisms bias genome scans for signatures of selection. *BMC Evol Biol.* 2012;12: 94.
4. Luikart G, England PR, Tallmon D, Jordan S, Taberlet P. The power and promise of population genomics: from genotyping to genome typing. *Nat Rev Genet.* 2003;4: 981–94.
5. Vandepitte K, Honnay O, Mergeay J, Breyne P, Roldán-Ruiz I, De Meyer T. SNP discovery using Paired-End RAD-tag sequencing on pooled genomic DNA of *Sisymbrium austriacum* (Brassicaceae). *Mol Ecol Resour.* 2013;13: 269–75.
6. Larson WA, Seeb LW, Everett M V., Waples RK, Templin WD, Seeb JE. Genotyping by sequencing resolves shallow population structure to inform conservation of *Chinook salmon* (*Oncorhynchus tshawytscha*). *Evol Appl.* 2014;7: 355–69.
7. Dou J, Li X, Fu Q, Jiao W, Li Y, Li T, et al. Evaluation of the 2b-RAD method for genomic selection in scallop breeding. *Sci Rep.* 2016;6: 19244.
8. Foll M, Gaggiotti O. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics.* 2008;180: 977–93.
9. Fischer MC, Foll M, Excoffier L, Heckel G. Enhanced AFLP genome scans detect local adaptation in high-altitude populations of a small rodent (*Microtus arvalis*). *Mol Ecol.* 2011;20: 1450–62.

10. Foll M, Fischer MC, Heckel G, Excoffier L. Estimating population structure from AFLP amplification intensity. *Mol Ecol*. 2010;19: 4638–47.
11. Excoffier L, Lischer HEL. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour*. 2010;10: 564–7.
12. Beaumont MA, Nichols RA. Evaluating Loci for Use in the Genetic Analysis of Population Structure. *Proc R Soc London B Biol Sci*. 1996;263: 1619–26.
13. Caye K, Deist TM, Martins H, Michel O, François O. TESS3: Fast inference of spatial population structure and genome scans for selection. *Mol Ecol Resour*. 2015; 540–8.
14. Antao T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G. LOSITAN: A workbench to detect molecular adaptation based on a  $F_{st}$  -outlier method. *BMC Bioinformatics*. 2008;9: 1–5.
15. Jombart T, Ahmed I. adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*. 2011;27: 3070–1.
16. Frichot E, Mathieu F, Trouillon T, Bouchard G, François O. Fast and efficient estimation of individual ancestry coefficients. *Genetics*. 2014;196: 973–83.
17. Stamatakis A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014;30: 1312–3.
18. Pickrell JK, Pritchard JK. Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. *PLoS Genet*. 2012;8.
19. Raghavan M, Skoglund P, Graf KE, Metspalu M, Albrechtsen A, Moltke I, et al. Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature*, 2014; 7481: 87-91.

Table 2.1 Sample list

Sample name	ID	Sampling location	remarks
sandan102	a1-102	Sandanpeki Shirahama-cho, Wakayama	TSUKUBA
sandan103	a1-103	Sandanpeki Shirahama-cho, Wakayama	TSUKUBA
shionomisaki106	a1-106	Shionomisaki Kusimoto-cho, Wakayama	TSUKUBA
shionomisaki108	a1-108	Shionomisaki Kusimoto-cho, Wakayama	TSUKUBA
kantori117	a1-117	Kantori-saki Taichi-cho, Wakayama	TSUKUBA
esan132	a1-132	Esan Todohokke-mura, Hokkaido	TSUKUBA
tsumeki14	a1-14	Tsumekisaki Shimoda-shi, Shizuoka	TSUKUBA
tatimati145	a1-145	Tachimachimisaki Hakodate-shi, Hokkaido	TSUKUBA
jougasaki15	a1-15	Jogasakikaigan Ito-shi, Shizuoka	TSUKUBA
tanesashi153	a1-153	Hukuura-cho Higashitugaru-gun, Aomori	TSUKUBA
tanesashi155	a1-155	Hukuura-cho Higashitugaru-gun, Aomori	TSUKUBA
jougasaki32	a1-32	Jogasakikaigan Ito-shi, Shizuoka	TSUKUBA
jougasaki37	a1-37	Jogasakikaigan Ito-shi, Shizuoka	TSUKUBA
usuki09	a1-9	Usuki Miyake-jima, Tokyo	TSUKUBA
esan1-35	a1-e35	Esan Todohokke-mura, Hokkaido	used in chapter 1
esan1-37	a1-e37	Esan Todohokke-mura, Hokkaido	used in chapter 1
esan1-41	a1-e41	Esan Todohokke-mura, Hokkaido	used in chapter 1
esan1-05	a1-e5	Esan Todohokke-mura, Hokkaido	used in chapter 1
esan1-07	a1-e7	Esan Todohokke-mura, Hokkaido	used in chapter 1
esan2-11	a1-ee11	Esan Todohokke-mura, Hokkaido	used in chapter 1
esan2-21	a1-ee21	Esan Todohokke-mura, Hokkaido	used in chapter 1
esan2-04	a1-ee4	Esan Todohokke-mura, Hokkaido	used in chapter 1
esan2-08	a1-ee8	Esan Todohokke-mura, Hokkaido	used in chapter 1
esan2-09	a1-ee9	Esan Todohokke-mura, Hokkaido	used in chapter 1

izu_01	a1-j01	Jogasakikaigan Ito-shi, Shizuoka	
izu_02	a1-j02	Jogasakikaigan Ito-shi, Shizuoka	
izu_03	a1-j03	Jogasakikaigan Ito-shi, Shizuoka	
izu_04	a1-j04	Jogasakikaigan Ito-shi, Shizuoka	
izu_05	a1-j05	Jogasakikaigan Ito-shi, Shizuoka	
izu_06	a1-j06	Jogasakikaigan Ito-shi, Shizuoka	
izu_07	a1-j07	Jogasakikaigan Ito-shi, Shizuoka	
izu_08	a1-j08	Jogasakikaigan Ito-shi, Shizuoka	
izu_09	a1-j09	Jogasakikaigan Ito-shi, Shizuoka	
izu_10	a1-j10	Jogasakikaigan Ito-shi, Shizuoka	
izu_11	a1-j11	Jogasakikaigan Ito-shi, Shizuoka	
izu_12	a1-j12	Jogasakikaigan Ito-shi, Shizuoka	
a1-sado10	a1-sado10	Senkaku-wan Sadogasima, Niigata	used in chapter 1
a1-sado13	a1-sado13	Senkaku-wan Sadogasima, Niigata	used in chapter 1
a1-sado14	a1-sado14	Senkaku-wan Sadogasima, Niigata	used in chapter 1
a1-sado19	a1-sado19	Senkaku-wan Sadogasima, Niigata	used in chapter 1
a1-sado25	a1-sado25	Senkaku-wan Sadogasima, Niigata	used in chapter 1
a1-sado29	a1-sado29	Senkaku-wan Sadogasima, Niigata	used in chapter 1
a1-sado31	a1-sado31	Senkaku-wan Sadogasima, Niigata	used in chapter 1
a1-sado8	a1-sado8	Senkaku-wan Sadogasima, Niigata	used in chapter 1
a1-san5	a1-san5	Sandanpeki Shirahama-cho, Wakayama	used in chapter 1
a1-san12	a1-san12	Sandanpeki Shirahama-cho, Wakayama	used in chapter 1
a1-san16	a1-san16	Sandanpeki Shirahama-cho, Wakayama	used in chapter 1
a1-san18	a1-san18	Sandanpeki Shirahama-cho, Wakayama	used in chapter 1
a1-san20	a1-san20	Sandanpeki Shirahama-cho, Wakayama	used in chapter 1
a1-san21	a1-san21	Sandanpeki Shirahama-cho, Wakayama	used in chapter 1
a1-san24	a1-san24	Sandanpeki Shirahama-cho, Wakayama	used in chapter 1

a1-san7	a1-san7	Sandanpeki Shirahama-cho, Wakayama	used in chapter 1
satamisaki13	a2-13	Satamisaki Minamiookuma-cho, Kagoshima	TSUKUBA
ogas15	a2-15	Sakaiura Chichijima-retto, Tokyo	TSUKUBA
ogas16	a2-16	Hatsune Chichijima-retto, Tokyo	TSUKUBA
ogas18	a2-18	Sakaiura Chichijima-retto, Tokyo	TSUKUBA
ogas19	a2-19	Mt.mikazuki Chichijima-retto, Tokyo	TSUKUBA
ogas02	a2-2	Mt.mikazuki Chichijima-retto, Tokyo	TSUKUBA
makurazaki29	a2-29	yamatatugami makurazaki kaoshima	TSUKUBA
kunigamiheira49	a2-49	Kushibaru Nagoshi, Okinawa	TSUKUBA
ogas08	a2-8	Hatsune Chichijima-retto, Tokyo	TSUKUBA
nomo21	a2-gs21	Nomozaki Nomozaki-cho, Nagasaki	
nomo22	a2-gs22	Nomozaki Nomozaki-cho, Nagasaki	
nomo23	a2-gs23	Nomozaki Nomozaki-cho, Nagasaki	
nomo05	a2-gs5	Nomozaki Nomozaki-cho, Nagasaki	
naka01	a2-ka1	Nomozaki Nomozaki-cho, Nagasaki	
naka02	a2-ka2	Nomozaki Nomozaki-cho, Nagasaki	
naka03	a2-ka3	Nomozaki Nomozaki-cho, Nagasaki	
naka04	a2-ka4	Nomozaki Nomozaki-cho, Nagasaki	
naka05	a2-ka5	Nomozaki Nomozaki-cho, Nagasaki	
naka06	a2-ka6	Nomozaki Nomozaki-cho, Nagasaki	
naka07	a2-ka7	Nomozaki Nomozaki-cho, Nagasaki	
naka08	a2-KA8	Nomozaki Nomozaki-cho, Nagasaki	
kasa01	a2-ks1	Kasasa-cho Satsuma-shi, Kagoshima	
kasa02	a2-ks2	Kasasa-cho Satsuma-shi, Kagoshima	
kasa03	a2-ks3	Kasasa-cho Satsuma-shi, Kagoshima	
kasa04	a2-ks4	Kasasa-cho Satsuma-shi, Kagoshima	
kasa05	a2-ks5	Kasasa-cho Satsuma-shi, Kagoshima	

kasa06	a2-ks6	Kasasa-cho Satsuma-shi, Kagoshima	
tuno03	a2-t3	Tunosima Shimonoseki-shi, Yamaguchi	
tuno04	a2-t4	Tunosima Shimonoseki-shi, Yamaguchi	
tuno05	a2-t5	Tunosima Shimonoseki-shi, Yamaguchi	
tuno07	a2-t7	Tunosima Shimonoseki-shi, Yamaguchi	
tuno08	a2-t8	Tunosima Shimonoseki-shi, Yamaguchi	
gm01	gm01	A1-55xA2-2	gametophyte
gm02	gm02	A1-55xA2-3	gametophyte
gm03	gm03	A1-55xA2-4	gametophyte
gm04	gm04	A1-55xA2-5	gametophyte
gm05	gm05	A1-55xA2-6	gametophyte
gm06	gm06	A1-55xA2-7	gametophyte
gm07	gm07	A1-55xA2-8	gametophyte
gm08	gm08	A1-55xA2-9	gametophyte
gm09	gm09	A1-55xA2-10	gametophyte
gm10	gm10	A1-55xA2-11	gametophyte
gm11	gm11	A1-55xA2-12	gametophyte
gm12	gm12	A1-55xA2-13	gametophyte

TSUKUBA: Sampled in TSUKUBA Botanical garden

Table 2.2  $F_{IS}$  and  $He$  values of each populations.

	ESAN	IZU	SAND	SADO	KASA	NAKA	NOMO	TUNO	OGAS
FIS	0.4799	0.3082	0.407 (0.491)	-0.682 (-0.0612)	-0.1099	0.2336	-0.8785	-0.4282	0.430
He	0.016	0.083	0.022 (0.125)	0.185 (0.241)	0.265	0.229	0.219	0.250	0.181

Population genetic parameters calculated using all samples including the admixed individuals revealed by the cluster analysis are shown in parentheses.

Table 2.3 FST values between pairs of sexual diploid populations of *Cyrtomium falcatum*.

	ESAN	IZU	SAND	SADO	KASA	NAKA	NOMO	TUNO	OGAS
ESAN	0								
IZU	0.54248	0							
SAND	0.61193	0.46056	0						
SADO	0.65519	0.51166	0.36997	0					
KASA	0.63556	0.50354	0.35026	0.04186	0				
NAKA	0.70144	0.60191	0.46524	0.14822	0.09287	0			
NOMO	0.78943	0.65488	0.49805	0.23091	0.13065	0.23263	0		
TUNO	0.73325	0.60992	0.4486	0.1567	0.07064	0.16633	-0.0948	0	
OGAS	0.77348	0.64064	0.5192	0.3055	0.26116	0.32517	0.40395	0.34627	0

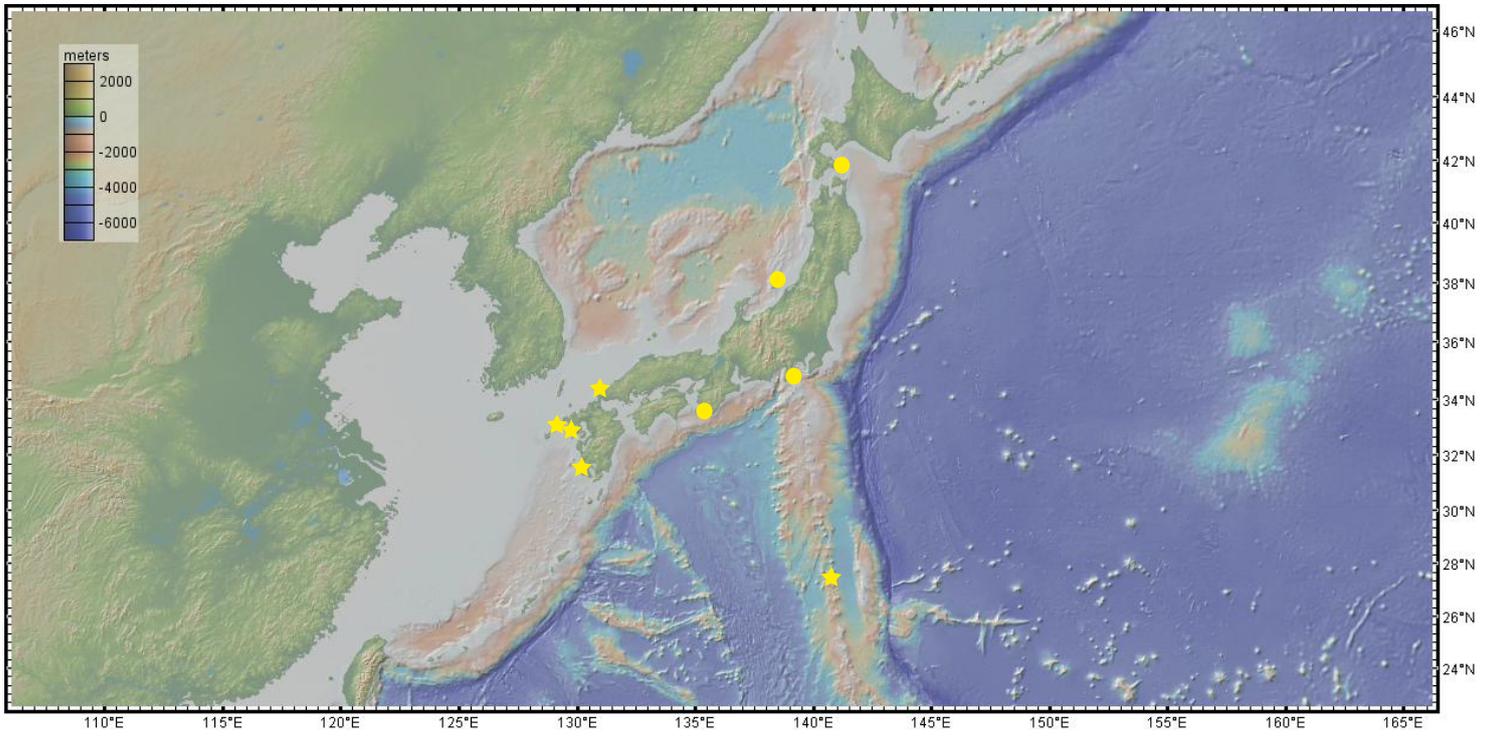


Fig 2.1 Stars indicate sampling location of populations of *C. falcatum* subsp. *australe*. Circles indicate populations sampled in chapter 1, those populations were used for analyses of chapter 2.

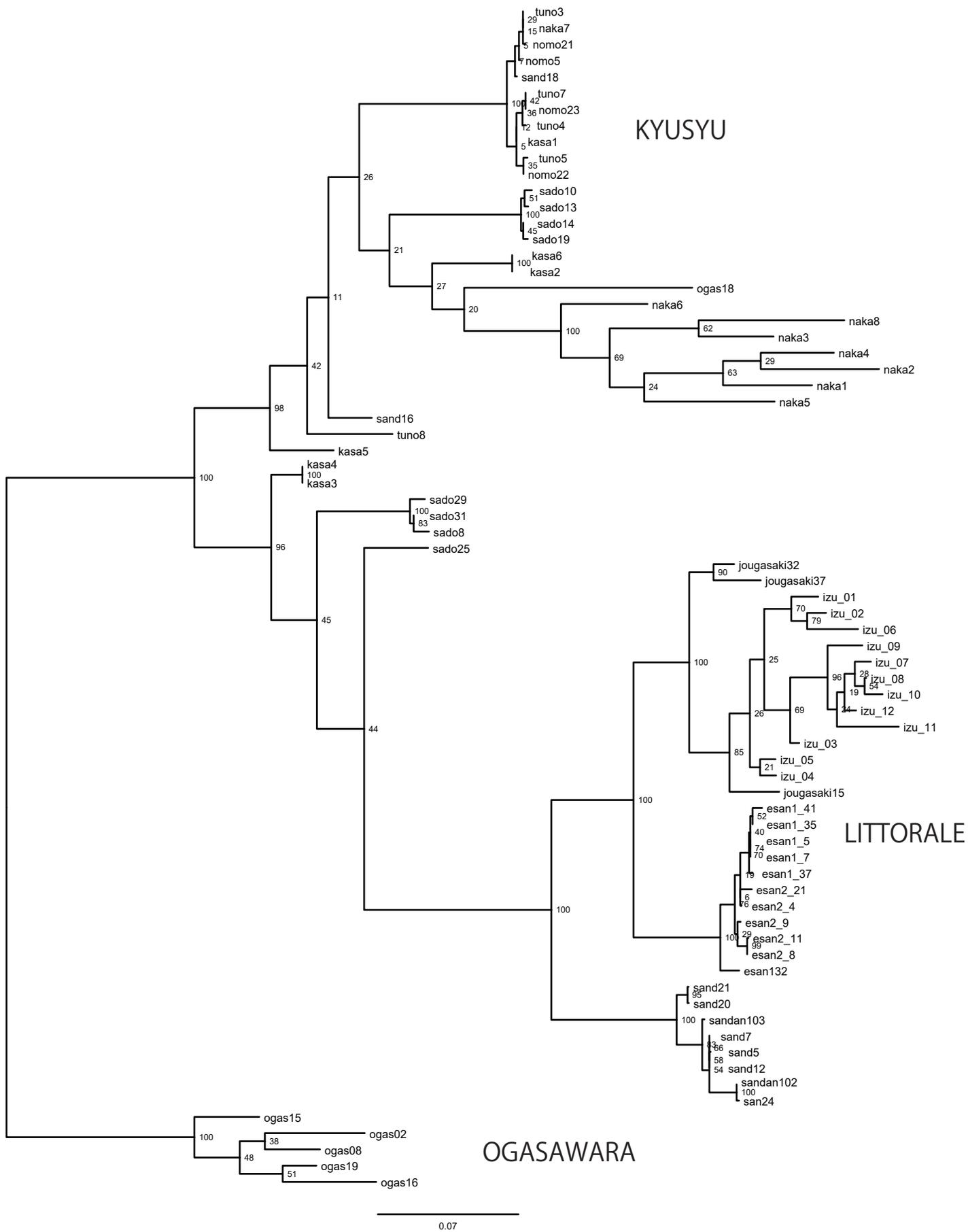


Fig. 2.2 Unrooted ML tree constructed by RAxML

3 clusters were recognized from tree topology and sampling location.

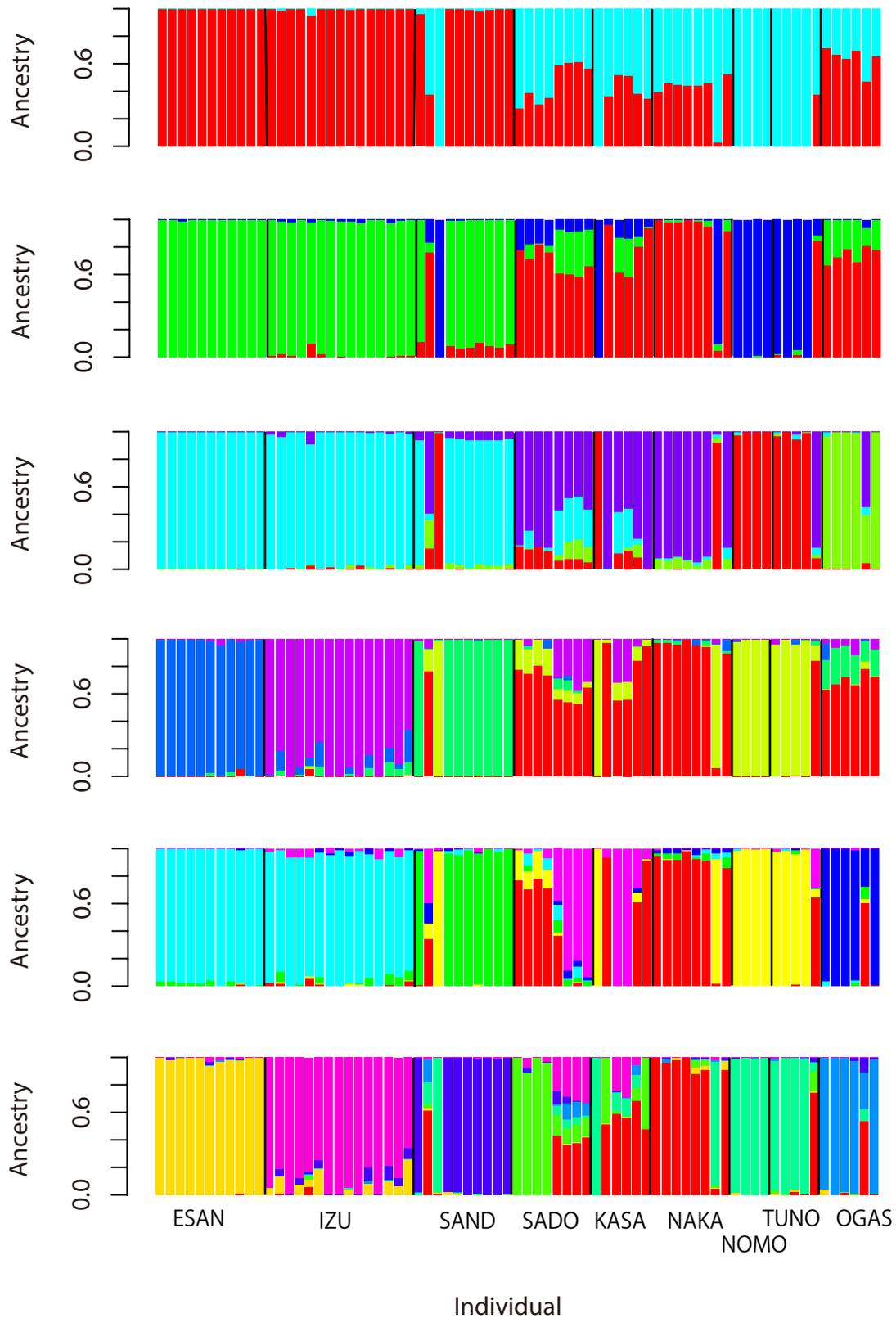


Fig. 2.3 The proportion of the membership coefficient of 73 individuals in 9 populations for each of the inferred clusters for  $K=2-7$  defined using sparse nonnegative matrix factorization algorithms in sNMF. The bar charts representing each individual are displayed horizontally and each narrow bar represents the result for an individual. Individuals were separated according to ancestry coefficients.

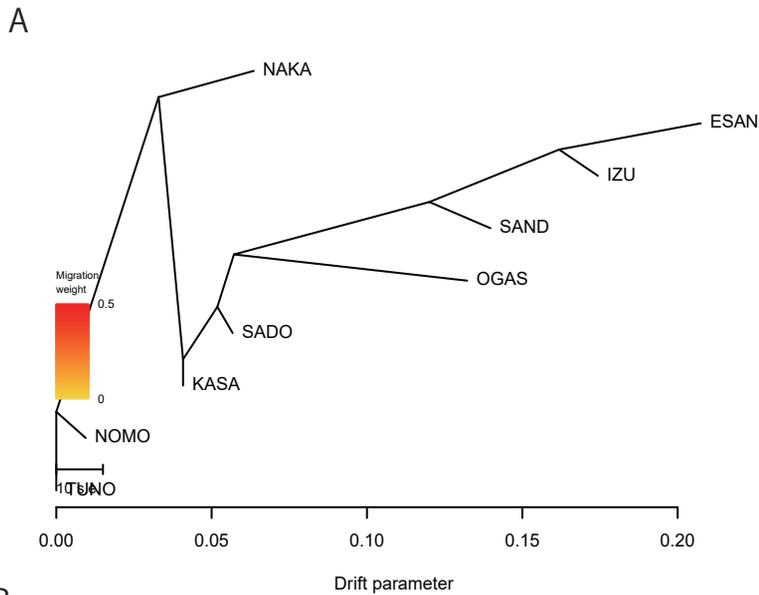
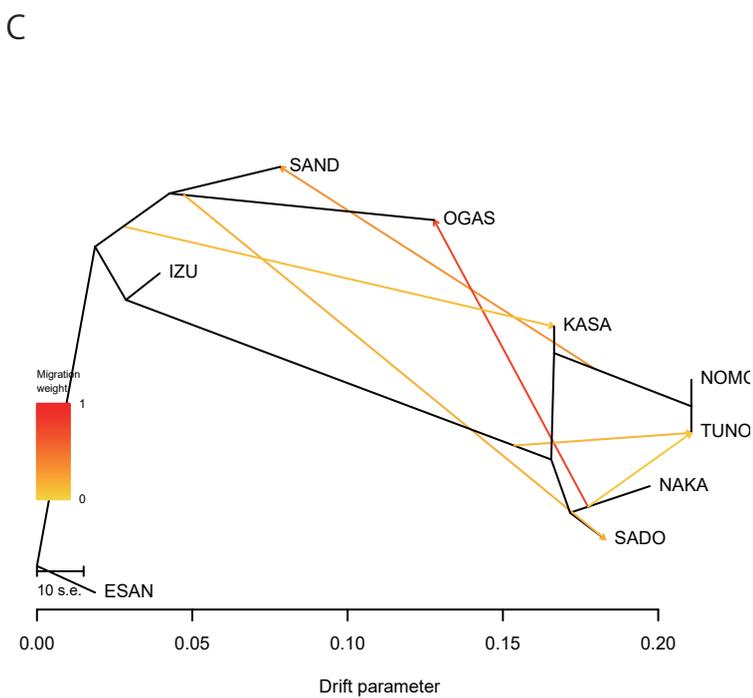
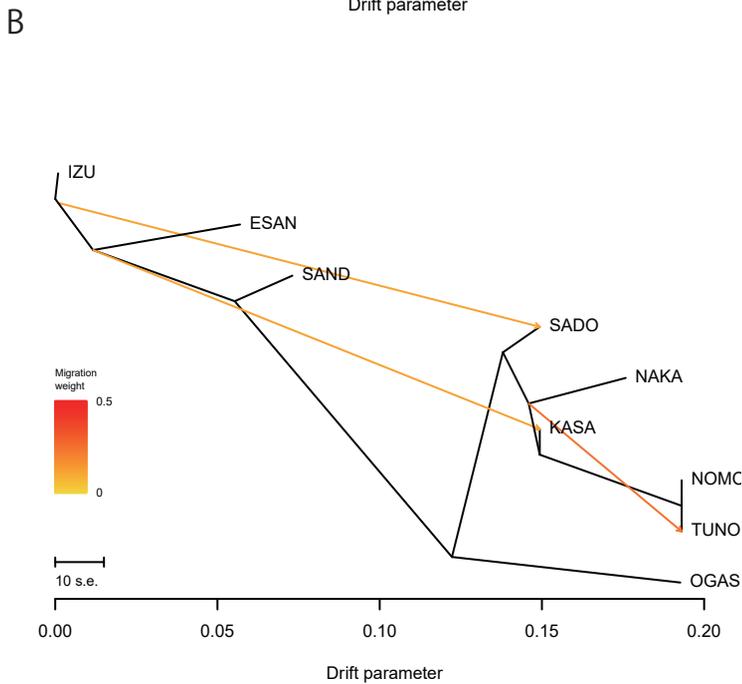


Fig. 2.4 Population tree with migration edges drawn by Treemix. The branch lengths are scaled to the amount of genetic drift between populations. Arrows indicate direction of migration and colors indicate migration weight. These trees were inferred under 0, 3 and 6 assumed migration.



## GENERAL DISCUSSION

### *Phylogenetic relationship among diploid populations of *Cyrtomium falcatum**

Both ML tree of individuals (Fig. 2.2) and population tree by Treemix suggested that *Cyrtomium falcatum* subsp. *littorale* except SADO and some individuals of SAND (defined as LITTORALE in chapter 2) form a well-supported group. Subspecies *littorale* is distinguished from subsp. *australe* by the characters of smaller blades, fewer pairs of pinnae, and grayish indusia without a blackish brown center (Matsumoto 2003). Based on these characters, SADO individuals are identified as subsp. *littorale* by Matsumoto (2003) and also in chapter 1. In this stage, therefore, we have no reliable morphological characters to identify the two subspecies. Possibly, ongoing and historical gene flow suggested by cluster analyses of SNPs (Fig. 2.3) and Treemix (Fig. 2.4) could have obscured the boundary of the two subspecies.

### *Evolution of mating system in diploid *Cyrtomium falcatum**

In chapter 1, it was shown that *C. falcatum* subsp. *littorale* except SADO, namely LITTORALE, has a mixed mating system based on the intermediate  $F_{IS}$  values (Table 1.2).  $F_{IS}$  value of SADO did not deviate from zero, suggesting that SADO was an outcrossing population. Additionally, *C. falcatum* subsp. *australe* have been considered to have outcrossing mating system, based on the S-type gametangium formation and low sporophyte formation rates in isolated gametophyte culture condition (Matsumoto 2003). Because LITTORALE was recognized as a well-supported group in ML tree, this suggests that transition from outcrossing to mixed mating would have occurred only once at the edge leading to LITTORALE in diploid populations of *C. falcatum*.

The M-type populations (ESAN, SAND, KANT) showed significantly lower

genetic diversities and effective population sizes than the S-type population in subsp. *littorale* in Chapter 1 using microsatellite markers (Table 1.2 , 1.3). The  $H_E$  values in Table 2.2 were calculated based on all samples included in the SNPs dataset. The ML tree (Fig. 2.2) and cluster analyses (Fig. 2.3) clearly indicated that some populations contained migrants and (or) admixed individuals. Therefore, it would be reasonable to recalculate gene diversities. The  $H_E$  values based on a new dataset excluding these putative migrants and(or) admixed individuals are as follows: ESAN (M-type, LITTORALE), 0.016; IZU (S-type, LITTORALE), 0.083; SAND (M-type, LITTORALE), 0.022; SADO1 (sado10, 13, 14, 19), 0.185; SADO2 (sado8, 25, 29, 31), 0.228; KASA (KYUSHU), 0.265; NAKA (KYUSHU): 0.229; NOMO (KYUSHU): 0.219, TUNO (KYUSHU): 0.250, OGAS (OGASAWARA): 0.181. Therefore, the expected heterozygosity ( $H_E$ ) values calculated by using SNPs supported the trends observed in microsatellite dataset. Additionally, the SNPs data indicated that genetic diversities of LITTORALE are lower than those of SADO, KYUSHU and OGASAWARA.

Matsumoto (2003) conducted gametophyte isolation experiments for subsp. *littorale* and subsp. *australe*. Gametophytes of subsp. *australe* can produce sporophyte only at low frequency mostly less than 50% when they are cultured under the condition isolated from other individuals. The low ability of sporophyte formation is usually considered to indicate high genetic load level (accumulation of recessive deleterious genes). In contrast to subsp. *australe*, the M-type individuals of subsp. *littorale* show nearly 100% sporophyte formation at isolation experiments. Although the S-type individuals of subsp. *littorale* show low levels of sporophyte formation just like subsp. *australe*, Matsumoto (2003) examined whether the low sporophyte formation in the S-type individuals of subsp. *littorale* is due to genetic load, or due to no chance of fertilization caused by its S-type sexual expression, by putting a vegetatively cloned gametophyte aside its original gametophyte. Interestingly, individuals of the

S-type collected from Jogasaki, Shizuoka (IZU, jogasaki and izu in this thesis) showed 100% sporophyte formation in this situation, suggesting low genetic load in the population. The results clearly show that LITTORARE group can produce sporophyte through intragametophytic selfing if they have chances of fertilization. Unfortunately, the test of sporophyte formation ability conducted for the S-type individuals of Jogasaki was not performed for individuals of SADO and subsp. *australe*[1]. The low genetic load in LITTORARE populations may be the genetic background that enable them to have a mixed mating system. Therefore, it may be important to test whether genetic loads of population of SADO and subsp. *australe* are higher than those of LITTORALE in future studies.

The low genetic load in LITTORALE populations might be caused by a historical bottleneck event that had occurred in the common ancestor of LITTORALE, and this could triggered the evolution of mixed mating in this lineage. This idea, although speculative, would also explain the reduced genetic diversities in LITTORALE populations. Historical demographic changes along phylogenetic tree should be examined to test this idea in future studies.

## REFERENCE

1. Matsumoto S. Species ecological study on reproductive systems and speciation of *Cyrtomium falcatum* complex (Dryopteridaceae) in Japanese archipelago. Ann Tsukuba Bot Gard. 2003;22: 1–141.

## ACKNOWLEDGEMENTS

We thank Dr. Tadashi Kajita, Dr. Koji Takayama, and Ms. Yoshimi Shinmura for advice in the development of microsatellite markers, Dr. Leanne Key Faulks for English editing, Ayumi Tezuka, Atsushi J. Nagano and Satoko Kondo for experiment of NGS and Mr. Nobuo Imai for assistance with sample collection. This work was supported by the Research Support Program Fund 2014 from Chiba University and a Grant-in-Aid (No. 15K07180 to Y. Watano) from the Ministry of Education, Science and Culture of Japan.

Table 1S1. Genetic diversity indices and inbreeding coefficient values for seven populations of diploid *Cyrtomium falcatum*.

Locus	$N_A$	Size range (bp)	M_type populations					S_type populations				
			ESAN 1 $n=41$	ESAN 2 $n=36$	SAND (SAN D#) $n=21$ ( $n=16$ )	KAN T $n=28$	Mean	IZU1 $n=42$	IZU2 $n=35$	SAD O (SAD O# $n=30$ ( $n=17$ )	Mean	
CFL-07 9	14	316-340	$N_A$	3	2	6 (3)	5	4.000	5	3	4 (2)	4.000
			$A_R$	2.075	1.754	5.642 (1.50)	4.536	3.502	3.698	3.000	3.850 (1.993 )	3.516
			$H_E$	0.073	0.056	0.364 (0.118)	0.417	0.228	0.568	0.589	0.600 (0.484 )	0.586
			$F_{IS}$	0.664	<b>1.000</b>	<b>0.451</b> (0.500)	<b>0.556</b>	0.668	0.142	<b>0.382</b>	-0.353 (-0.68 4)	0.171
CFL-C3 2	9	181-202	$N_A$	3	3	5 (3)	2	3.25	8	6	5 (2)	6.333
			$A_R$	2.352	2.566	4.681 (1.5)	2.000	2.900	6.623	5.883	4.497 (1.250 )	5.668
			$H_E$	0.139	0.292	0.361 (0.118)	0.300	0.273	0.821	0.809	0.599 (0.060 1)	0.743
			$F_{IS}$	0.123	<b>0.886</b>	0.584 (0.500)	<b>1.000</b>	0.648	<b>0.258</b>	0.143	<b>0.236</b> ( <b>0.000</b> )	0.212
CFL-Z0	7	227-25	$N_A$	3	1	1	1	1.500	3	2	6	3.667

3		3				(1)				(6)		
			$A_R$	2.450	1.000	1.000	1.000	1.363	2.649	2.000	5.005	3.218
						(1.000)					(3.113	)
			$H_E$	0.109	0.000	0.000	0.000	0.027	0.498	0.508	0.701	0.569
						(0.000)					(0.626	)
			$F_{IS}$	0.489	-	-	-	0.489	<b>0.618</b>	0.262	0.213	0.364
						(-)					(-0.48	0)
CFL-B0	14	130-15	$N_A$	1	4	5	1	2.75	4	2	8	4.67
2		5				(3)					(5)	
			$A_R$	1.000	3.400	4.990	1.000	2.598	3.108	1.754	6.889	3.917
						(2.180)					(3.476	)
			$H_E$	0.000	0.444	0.664	0.000	0.277	0.184	0.059	0.831	0.358
						(0.420)					(0.706	)
			$F_{IS}$	-	<b>0.937</b>	<b>0.570</b>	-	0.754	0.207	1.000	-0.025	0.394
						( <b>0.860</b> )					(-0.38	4)
CFL-B1	5	172-18	$N_A$	3	2	4	3	3.00	2	1	3	2.00
2		4				(4)					(3)	
			$A_R$	2.712	1.725	4.000	2.607	2.761	1.794	1.000	3.000	1.931
						(2.632)					(2.889	)
			$H_E$	0.167	0.056	0.526	0.476	0.306	0.070	0.000	0.654	0.241
						(0.435)					(0.650	)
			$F_{IS}$	<b>0.550</b>	1.000	<b>0.888</b>	<b>0.850</b>	0.822	0.661	-	<b>0.631</b>	0.646
						( <b>0.836</b> )					(0.656	)
CFL-B1	8	237-25	$N_A$	1	1	3	3	2.00	3	4	2	3.00
3		1				(3)					(2)	
			$A_R$	1.000	1.000	3.000	2.996	1.999	2.434	3.217	2.000	2.550
						(2.819)					(1.878	)

			$H_E$	0.000	0.000	0.663	0.389	0.263	0.252	0.166	0.384	0.267
						(0.605)					(0.327	)
			$F_{IS}$	-	-	<b>0.774</b>	0.111	0.443	<b>0.898</b>	<b>0.655</b>	0.392	0.648
						( <b>0.793</b> )					(0.484	)
CFL-B1	2	187-18	$N_A$	2	2	2	1	1.75	2	2	2	2.00
6		9				(2)					(2)	
			$A_R$	1.937	1.977	1.999	1.000	1.728	2.000	2.000	1.969	1.990
						(1.444)					(1.566	)
			$H_E$	0.117	0.143	0.179	0.000	0.110	0.471	0.321	0.126	0.306
						(0.116)					(0.161	)
			$F_{IS}$	<b>0.792</b>	0.364	0.467	-	0.541	0.276	-0.038	-0.055	0.061
						(1.000)					(-0.06	7)
CFL-B1	5	275-28	$N_A$	1	2	4	2	2.25	3	2	3	2.67
7		9				(1)					(2)	
			$A_R$	1.000	2.000	3.698	1.992	2.173	2.380	1.985	2.998	2.454
						(1.000)					(1.989	)
			$H_E$	0.000	0.275	0.233	0.150	0.165	0.160	0.159	0.604	0.308
						(0.000)					(0.472	)
			$F_{IS}$	-	0.039	-0.073	-0.067	-0.03	-0.069	-0.079	0.117	-0.010
						(-)		4			(-0.09	1)
Mean	8	-	$N_A$	2.13	2.125	3.75	2.25	2.56	3.75	2.75	4.125	3.541
						(2.5)					(3)	
			$A_R$	1.816	1.928	3.626	2.141	2.378	3.086	2.467	3.776	3.110
						(1.759)						
			$H_E$	0.076	0.158	0.374	0.217	0.206	0.378	0.326	0.562	0.422
						0.227						
			$F_{IS}^*$	<b>0.501</b>	<b>0.671</b>	<b>0.602</b>	<b>0.560</b>	0.583	<b>0.340</b>	<b>0.220</b>	<b>0.157</b>	0.239
						( <b>0.794</b> )		5			(-0.10	0)

Bold type indicates significant  $F_{IS}$  values ( $p < 0.00089$ , simple Bonferroni correction for 5% level); \*, multi locus estimate.

$A_R$ , Allelic richness, and  $H_E$ , Expected heterozygosity, are indexes of genetic diversity. The  $F_{IS}$  is Wright's fixation index, that index reflects selfing rate.

#: Excluding subpopulation detected by STRUCTURE

Table1S2.Null allele frequencies at each locus estimated by INEST2

Locus			M_type population					S-type population			
			ESAN1	ESAN2	SAND (SAND1 ) n=21 (n=16)	KANT n=28	Average	IZU1 n=42	IZU2 n=35	SADO (SADO1 ) n=30 (n=17)	Average
CFL-079	Freq. of null alleles		0.00835	0.0668	0.0398 (0.014)	0.0572	0.0431	0.0653	0.0136	0.00940 (0.00633)	0.0324
	Low (95%)		0.000	0.000	0.000 (0.000)	0.000	0.000	0.000	0.000	0.000 (0.000)	0.000
	High (95%)		0.0475	0.179	0.139 (0.0726)	0.189	0.139	0.176	0.0835	0.0450 (0.036)	0.114
CFL-C32	Freq. of null alleles		0.00653	0.184	0.0486 (0.015)	0.253	0.120	0.0333	0.00813	0.0661 (0.0239)	0.0502
	Low (95%)		0.000	0.000	0.000 (0.000)	0.000	0.000	0.000	0.000	0.000 (0.000)	0.000
	High (95%)		0.0378	0.350	0.167 (0.0757)	0.466	0.253	0.0979	0.0480	0.247 (0.1416)	0.135
CFL-Z03	Freq. of null alleles		0.133	0.171	0.144 (0.1141)	0.199	0.161	0.0518	0.0524	0.0747 (0.00616)	0.0525
	Low (95%)		0.000	0.000	0.000 (0.000)	0.000	0.000	0.000	0.000	0.000 (0.000)	0.000
	High (95%)		0.3054	0.345	0.359 (0.3107)	0.418	0.356	0.176	0.217	0.195 (0.0365)	0.183
CFL-B02	Freq. of null allele		0.00738	0.0168	0.0181 (0.0179)	0.0200	0.0150	0.0563	0.0573	0.0114 (0.00718)	0.0343
	Low (95%)		0.000	0.000	0.000 (0.000)	0.000	0.000	0.000	0.000	0.000 (0.000)	0.000
	High (95%)		0.0428	0.0825	0.0782 (0.0914)	0.0986	0.0728	0.154	0.245	0.0522 (0.0382 )	0.124
CFL-B12	Freq. of null alleles		0.0403	0.0123	0.212 (0.1338)	0.0403	0.0760	0.0310	0.0114	0.246 (0.1931 )	0.0894
	Low (95%)		0.000	0.000	0.000 (0.000)	0.000	0.000	0.000	0.000	0.0977 (0.000)	0.0252
	High (95%)		0.138	0.0610	0.395 (0.3185)	0.186	0.194	0.121	0.0675	0.400 (0.3851 )	0.190

CFL-B13	Freq. of null alleles	0.00734	0.0107	0.0760 (0.0567)	0.0526	0.0366	0.209	0.0200	0.0671 (0.0468)	0.0949
	Low (95%)	0.000	0.000	0.000 (0.000)	0.000	0.000	0.000	0.000	0.000 (0.000)	0.000
	High (95%)	0.0362	0.0527	0.212 (0.182)	0.174	0.118	0.350	0.113	0.207 (0.208)	0.213
CFL-B16	Freq. of null alleles	0.0100	0.0762	0.0210 (0.0173)	0.019	0.0314	0.0523	0.0318	0.0285 (0.0172)	0.0372
	Low (95%)	0.000	0.000	0.000 (0.000)	0.000	0.000	0.000	0.000	0.000 (0.000)	0.000
	High (95%)	0.0586	0.204	0.0926 (0.0936)	0.0966	0.111	0.148	0.142	0.124 (0.098)	0.135
CFL-B17	Freq. of null alleles	0.0915	0.0408	0.0426 (0.0669)	0.0933	0.0673	0.0472	0.00758	0.0319 (0.0125)	0.0269
	Low (95%)	0.000	0.000	0.000 (0.000)	0.000	0.000	0.000	0.000	0.000 (0.000)	0.000
	High (95%)	0.244	0.125	0.131 (0.2102)	0.263	0.190	0.142	0.0469	0.123 (0.0735)	0.097

---

Table 1S3. Posterior probability of each scenario by DIY ABC

Scenario	1	2	3
ESAN1	0.5233 (0.5116-0.5351)	0.3384 (0.3274-0.3494)	0.1383 (0.1287-0.1478)
ESAN2	0.6435 (0.6377-0.6493)	0.3205 (0.3148-0.3262)	0.0360 (0.0327-0.0394)
SAND2	0.6069 (0.6009-0.6130)	0.3371 (0.3312-0.3429)	0.0560 (0.0521-0.0599)
KANT	0.6126 (0.6069-0.6184)	0.3355 (0.3299-0.3411)	0.0519 (0.0482-0.0555)
IZU1	0.4716 (0.4639-0.4794)	0.3398 (0.3324-0.3473)	0.1885 (0.1824-0.1946)
IZU2	0.6135 (0.6061-0.6209)	0.3467 (0.3394-0.3540)	0.0398 (0.0354-0.0442)
SADO2	0.7728 (0.7624-0.7832)	0.2003 (0.1902-0.2104)	0.0269 (0.0233-0.0305)

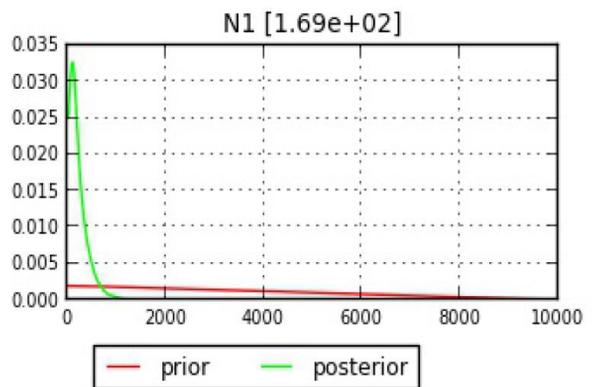
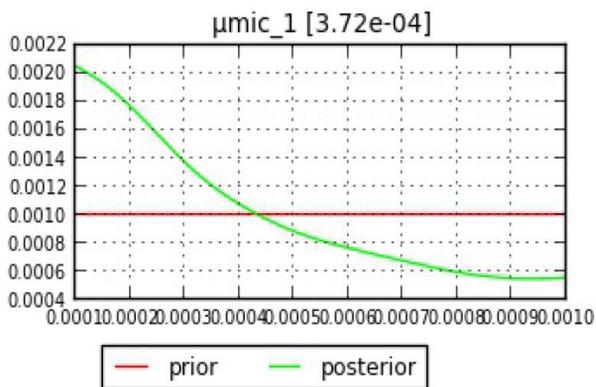
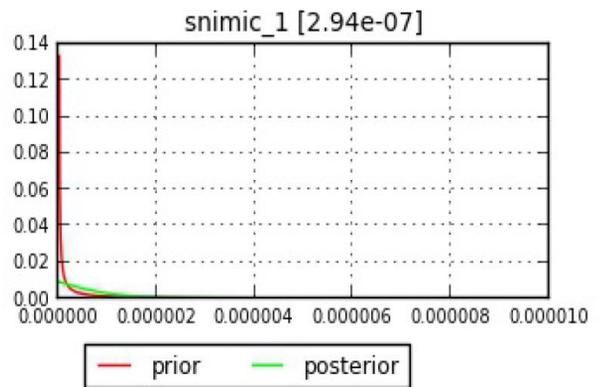
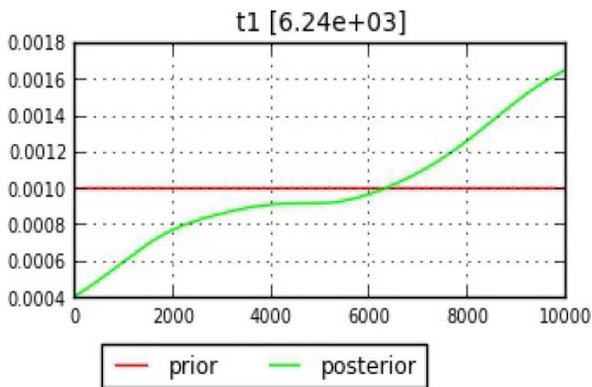
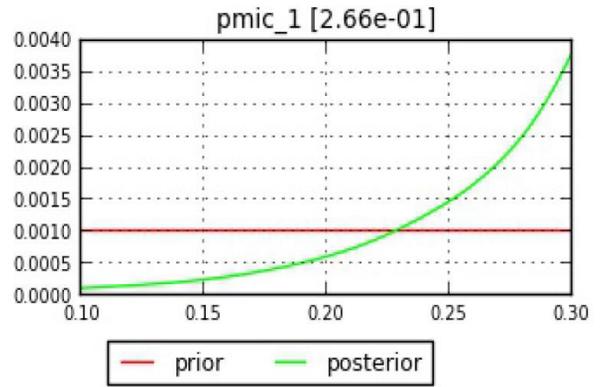
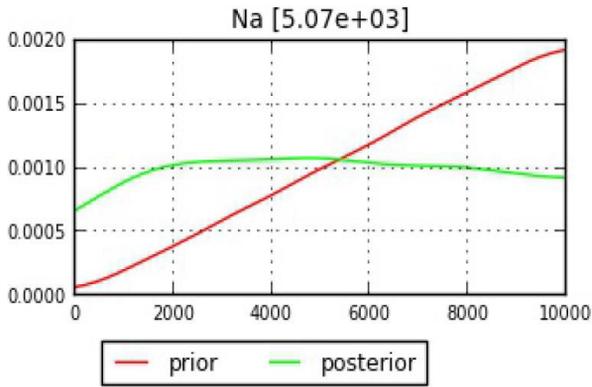
Table 1S4. Demographic paramters of the scenario 1 obtained by DIYABC

Parameter	quantile								
	mean	median	mode	2.50%	5%	25%	75%	95%	97.50%
ESAN1									
N1	230.00	169.00	77.90	39.90	50.00	95.60	303.00	607.00	742.00
t1	5930.00	6240.00	9950.00	796.00	1200.00	3600.00	8440.00	9740.00	9880.00
Na	5100.00	5070.00	5240.00	591.00	862.00	2750.00	7460.00	9470.00	9730.00
Mean mutation rate_SSR	4.33E-04	3.72E-04	1.08E-04	1.10E-04	1.20E-04	2.10E-04	6.27E-04	9.17E-04	9.61E-04
Mean P*	0.253	0.266	0.300	0.135	0.157	0.230	0.288	0.300	0.300
Mean mutation rate_SNI	1.33E-06	2.94E-07	1.01E-08	1.17E-08	1.41E-08	5.58E-08	1.56E-06	6.54E-06	8.02E-06
ESAN2									
N1	284	210	93.5	47.3	57.9	117	381	744	903
t1	5110	5060	1620	514	817	2600	7580	9550	9770
Na	5.13E+03	5.10E+03	2.30E+03	5.54E+02	8.39E+02	2.73E+03	7.49E+03	9.52E+03	9.76E+03
Mean mutation rate_SSR	4.03E-04	3.34E-04	1.05E-04	1.07E-04	1.14E-04	1.87E-04	5.80E-04	8.96E-04	9.44E-04
Mean P*	2.34E-01	2.44E-01	3.00E-01	1.19E-01	1.35E-01	2.01E-01	2.75E-01	2.97E-01	3.00E-01
Mean mutation rate_SNI	1.38E-06	3.02E-07	1.03E-08	1.20E-08	1.44E-08	5.46E-08	1.63E-06	6.93E-06	8.28E-06
SAND									
N1	405	299	150	68.6	84.7	171	530	1080	1300
t1	5240	5210	4570	620	944	2900	7590	9520	9770
Na	5160	5120	1950	604	889	2750	7550	9520	9770
Mean mutation rate_SSR	4.10E-04	3.44E-04	1.04E-04	1.07E-04	1.15E-04	1.91E-04	5.93E-04	9.02E-04	9.50E-04
Mean P*	0.217	0.224	0.279	0.11	0.12	0.175	0.263	0.293	0.297
Mean mutation rate_SNI	1.33E-06	2.97E-07	1.12E-08	1.19E-08	1.42E-08	5.62E-08	1.55E-06	6.59E-06	8.03E-06
KANT									
N1	432	328	135	78.4	94.5	188	578	1100	1330
t1	5150	5130	2380	571	843	2640	7680	9550	9780
Na	5230	5230	4700	631	939	2910	7630	9510	9750
Mean mutation rate_SSR	4.05E-04	3.37E-04	1.03E-04	1.07E-04	1.15E-04	1.91E-04	5.79E-04	9.01E-04	9.50E-04
Mean P*	0.23	0.239	0.3	0.118	0.132	0.195	0.273	0.296	0.299
Mean mutation rate_SNI	1.36E-06	2.89E-07	1.06E-08	1.16E-08	1.39E-08	5.41E-08	1.59E-06	6.83E-06	8.30E-06
IZU1									
N1	1090	846	461	238	281	504	1450	2700	3230
t1	5860	6030	9270	1050	1490	3760	8100	9640	9810
Na	5350	5310	4870	917	1230	3140	7600	9500	9740
Mean mutation rate_SSR	4.44E-04	3.90E-04	1.10E-04	1.10E-04	1.20E-04	2.14E-04	6.48E-04	9.20E-04	9.63E-04
Mean P*	0.208	0.211	0.226	0.108	0.116	0.164	0.255	0.291	0.296
Mean mutation rate_SNI	1.32E-06	2.83E-07	1.16E-08	1.18E-08	1.40E-08	5.27E-08	1.54E-06	6.68E-06	8.10E-06
IZU2									
N1	618	465	194	123	148	272	811	1600	1920
t1	5430	5460	7190	736	1110	3170	7730	9550	9790
Na	5150	5110	2520	730	1020	2790	7430	9460	9740
Mean mutation rate_SSR	4.09E-04	3.44E-04	1.00E-04	1.07E-04	1.14E-04	1.91E-04	5.91E-04	8.94E-04	9.47E-04
Mean P*	0.2	0.2	0.207	0.106	0.112	0.154	0.246	0.287	0.293
Mean mutation rate_SNI	1.31E-06	2.85E-07	1.05E-08	1.19E-08	1.40E-08	5.61E-08	1.51E-06	6.53E-06	8.00E-06
SADO									
N1	934	715	297	182	219	422	1230	2340	2800
t1	3650	2940	947	256	408	1350	5580	8870	9390
Na	6050	6260	8850	1280	1790	4080	8180	9670	9830
Mean mutation rate_SSR	3.60E-04	2.85E-04	1.07E-04	1.05E-04	1.10E-04	1.67E-04	5.03E-04	8.53E-04	9.15E-04
Mean P*	0.203	0.206	0.234	0.106	0.114	0.16	0.248	0.284	0.289
Mean mutation rate_SNI	1.28E-06	2.74E-07	1.09E-08	1.19E-08	1.42E-08	5.32E-08	1.47E-06	6.54E-06	8.02E-06

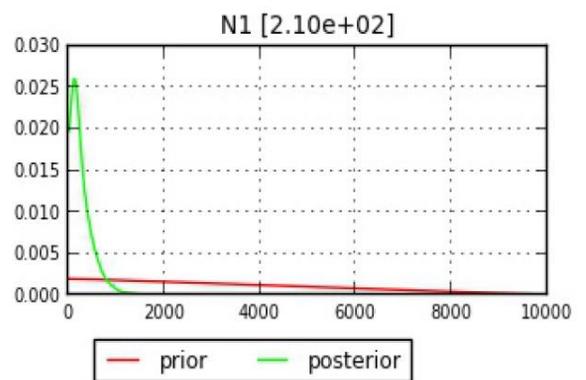
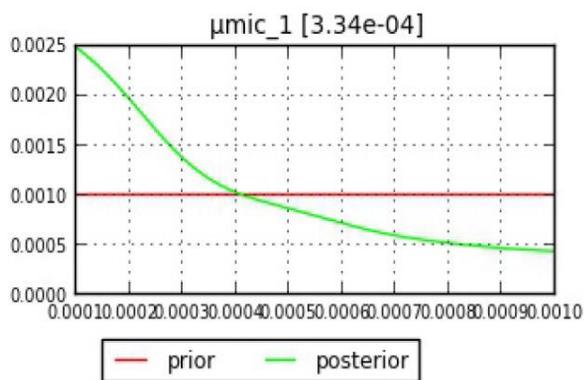
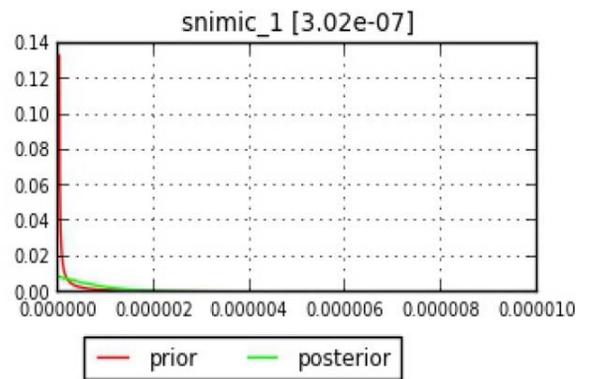
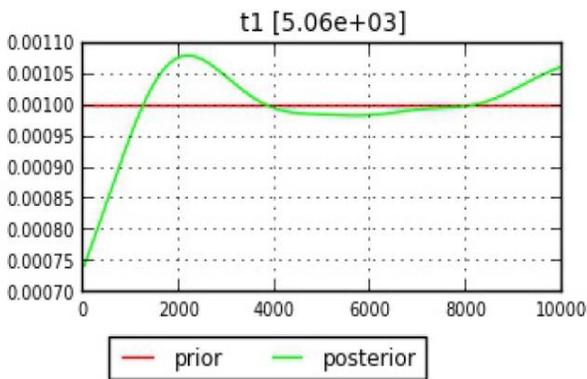
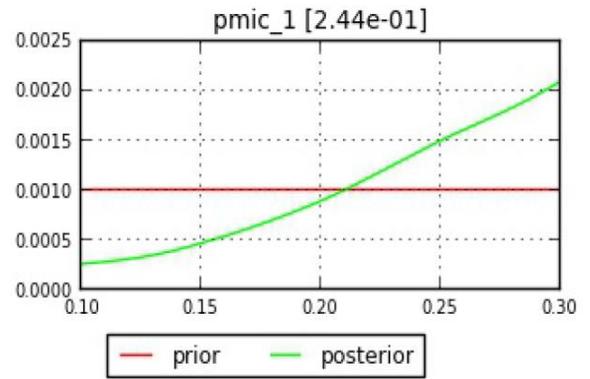
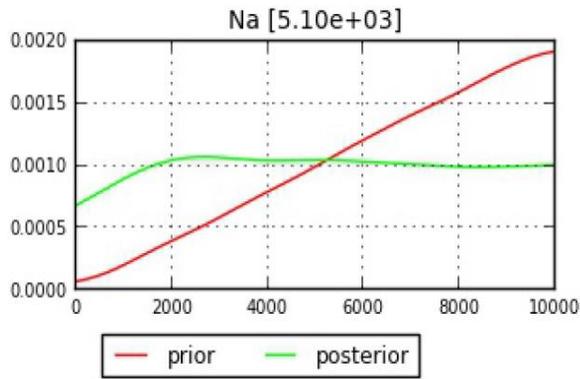
\*the parameter of the geometric distribution to generate multiple stepwise mutations

Fig. 1S1

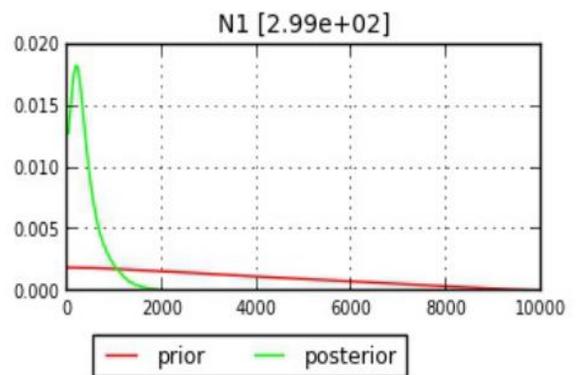
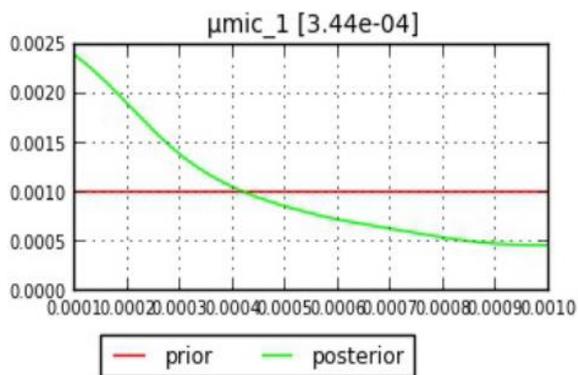
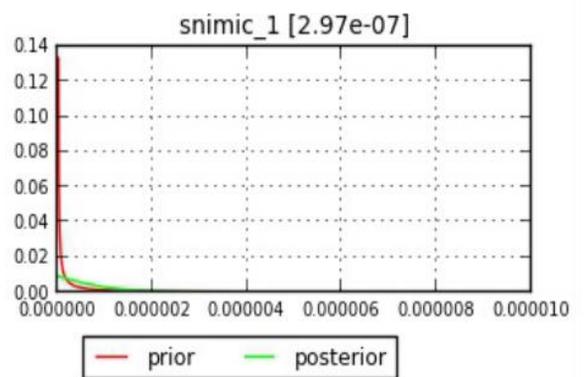
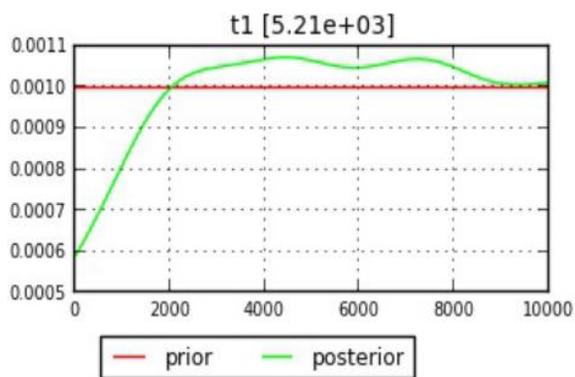
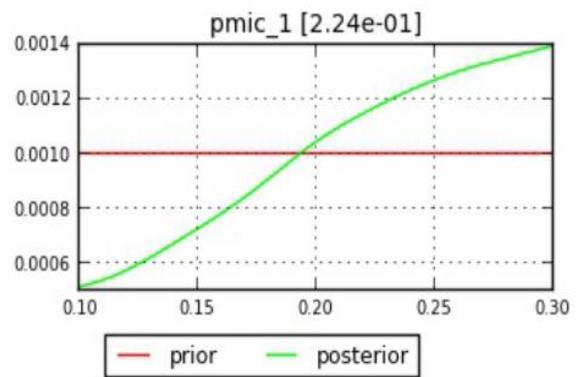
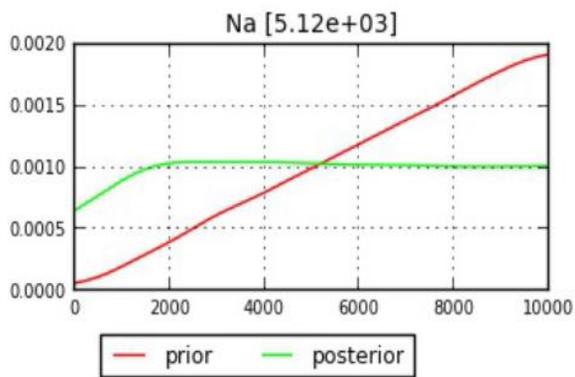
# ESAN1



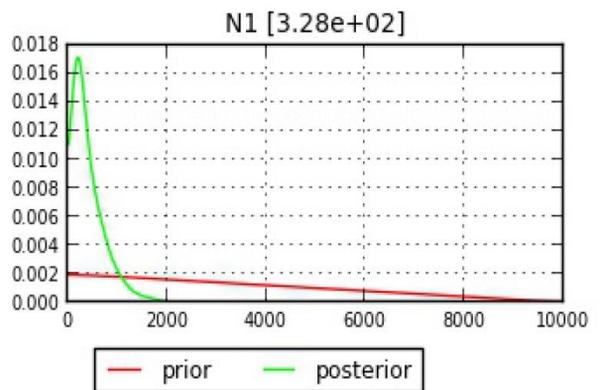
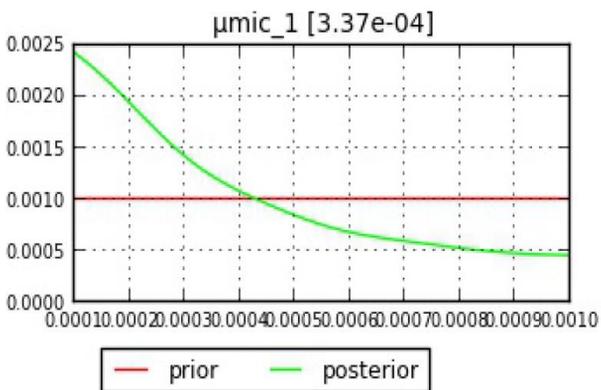
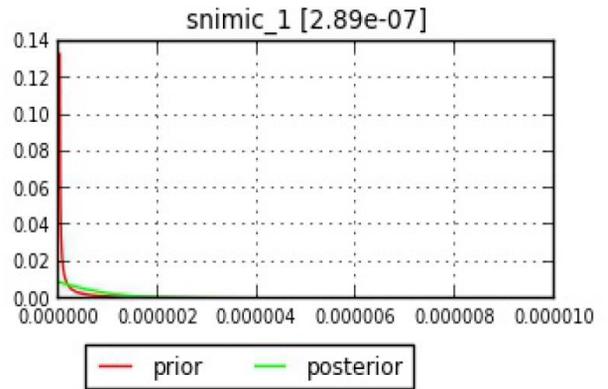
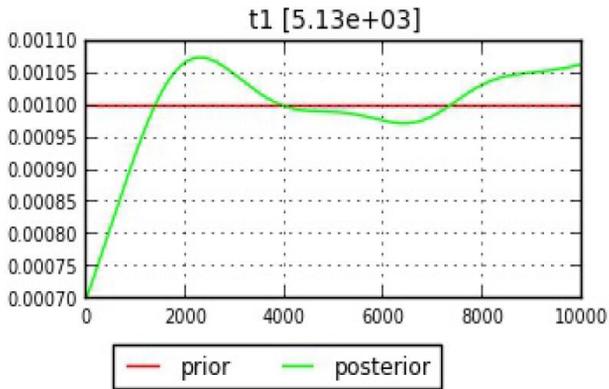
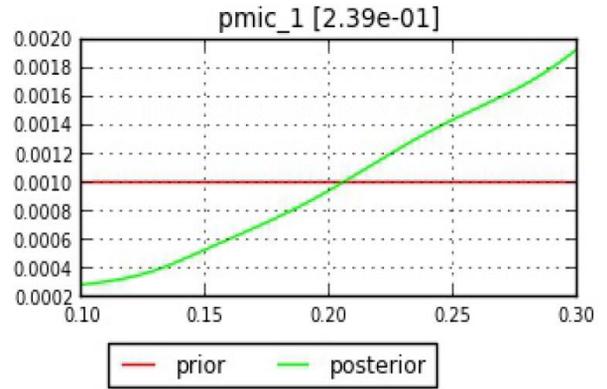
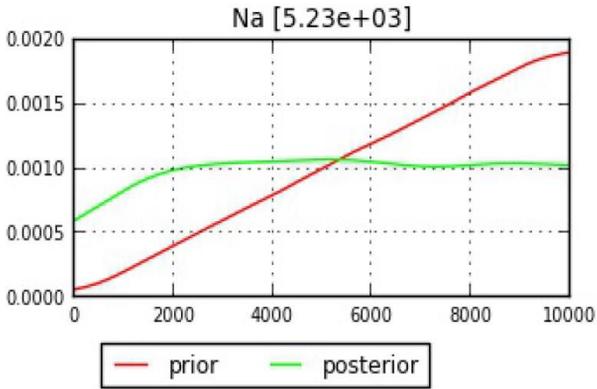
# ESAN2



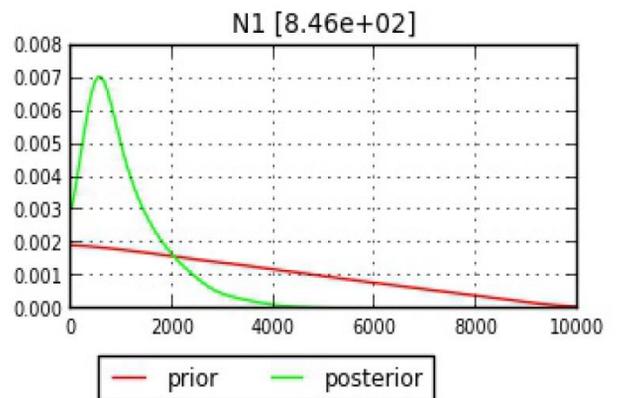
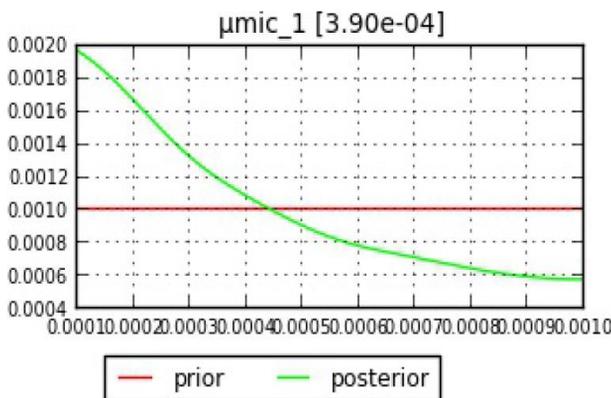
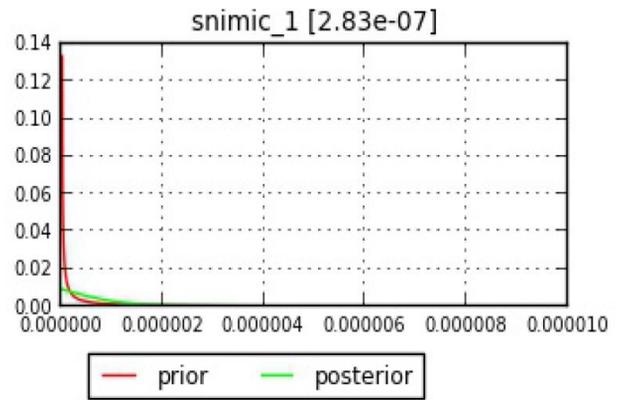
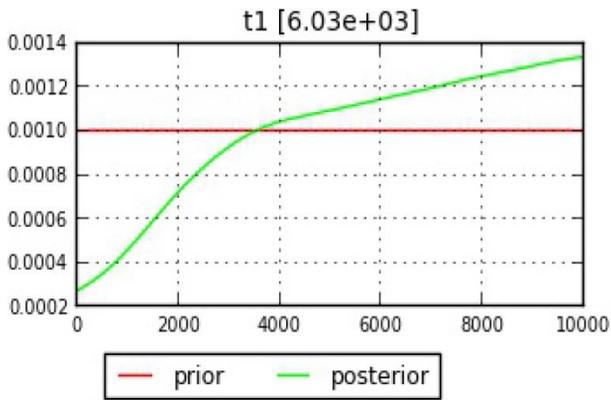
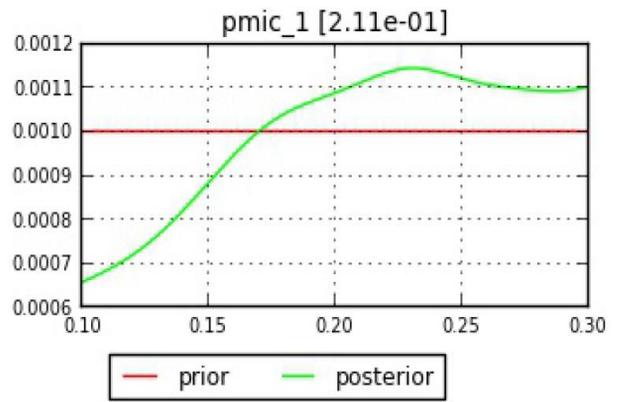
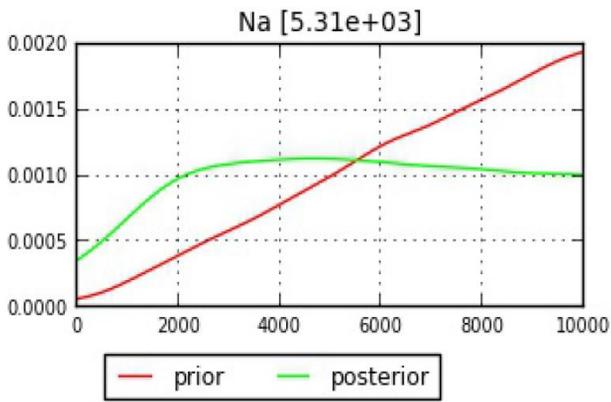
# SAND1

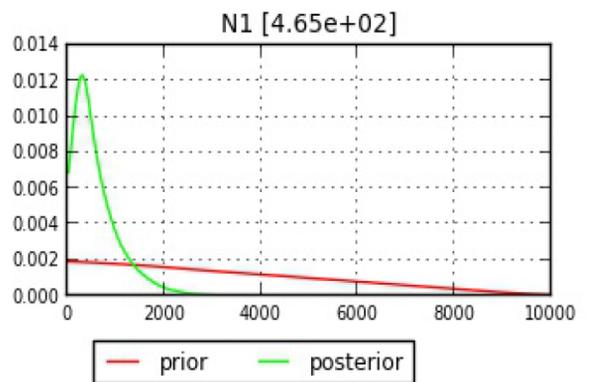
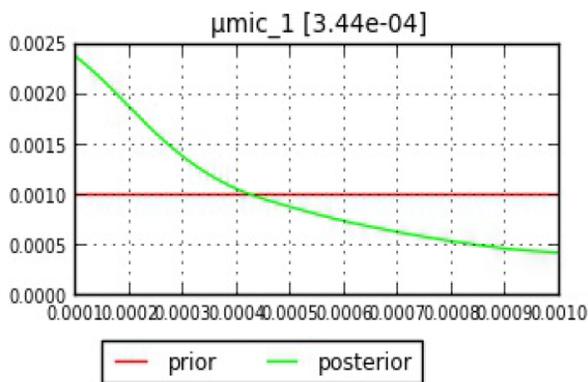
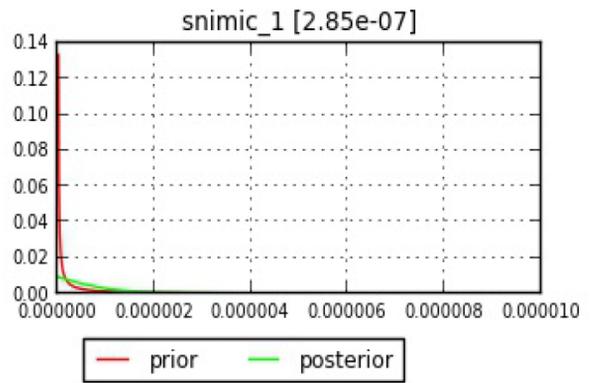
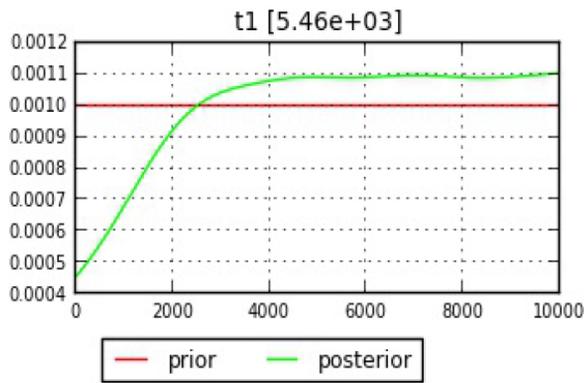
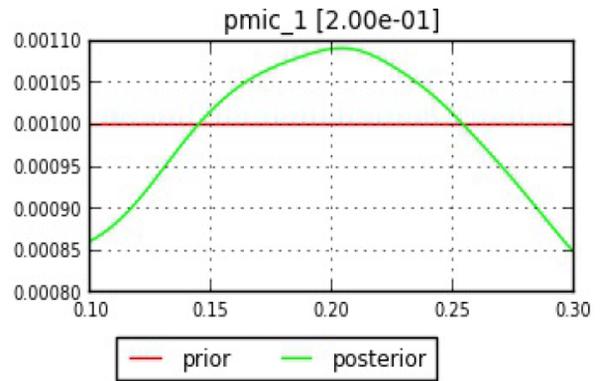
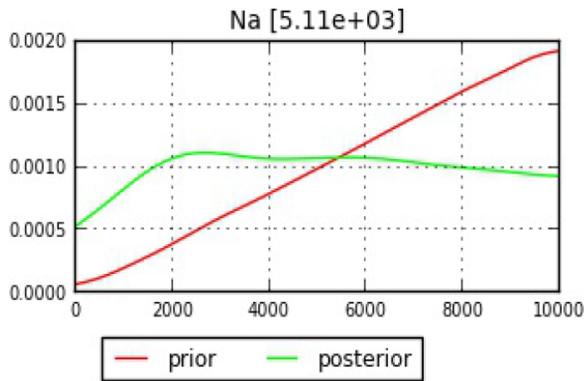


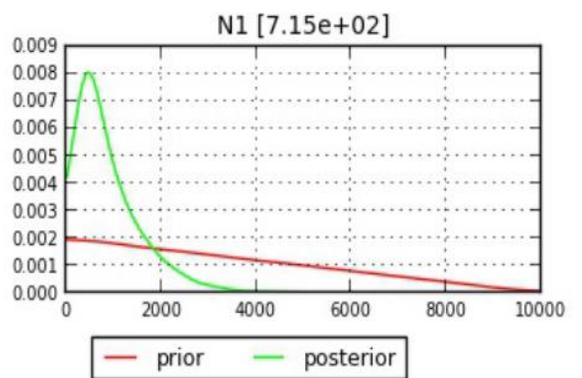
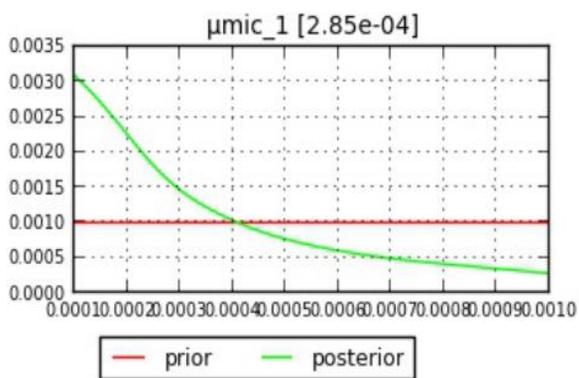
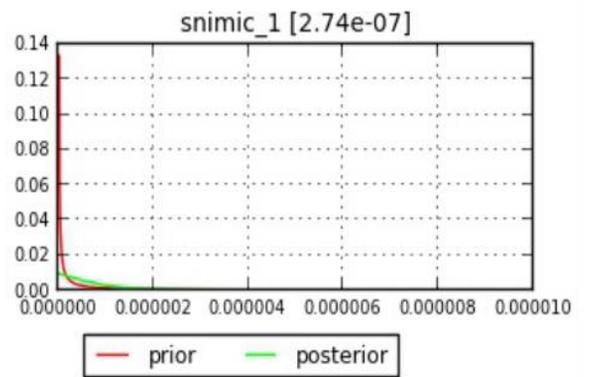
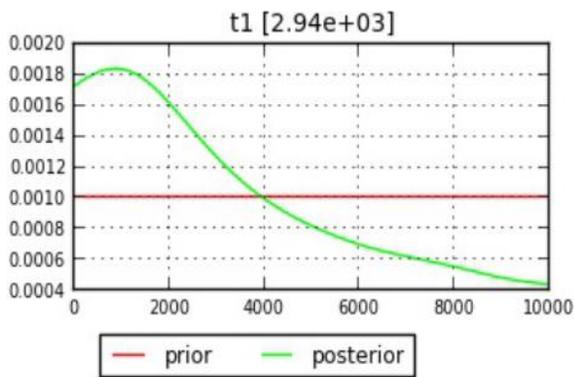
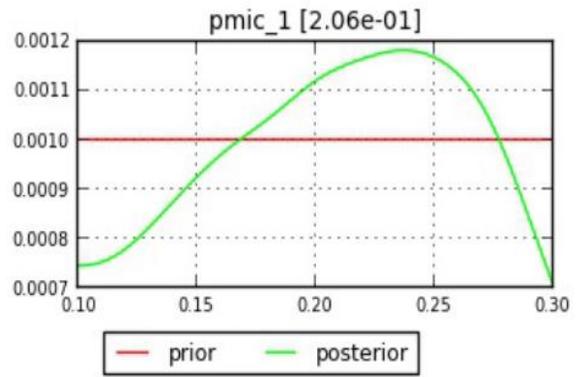
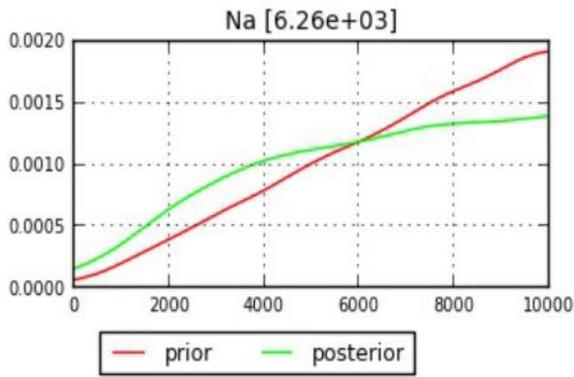
# KANT



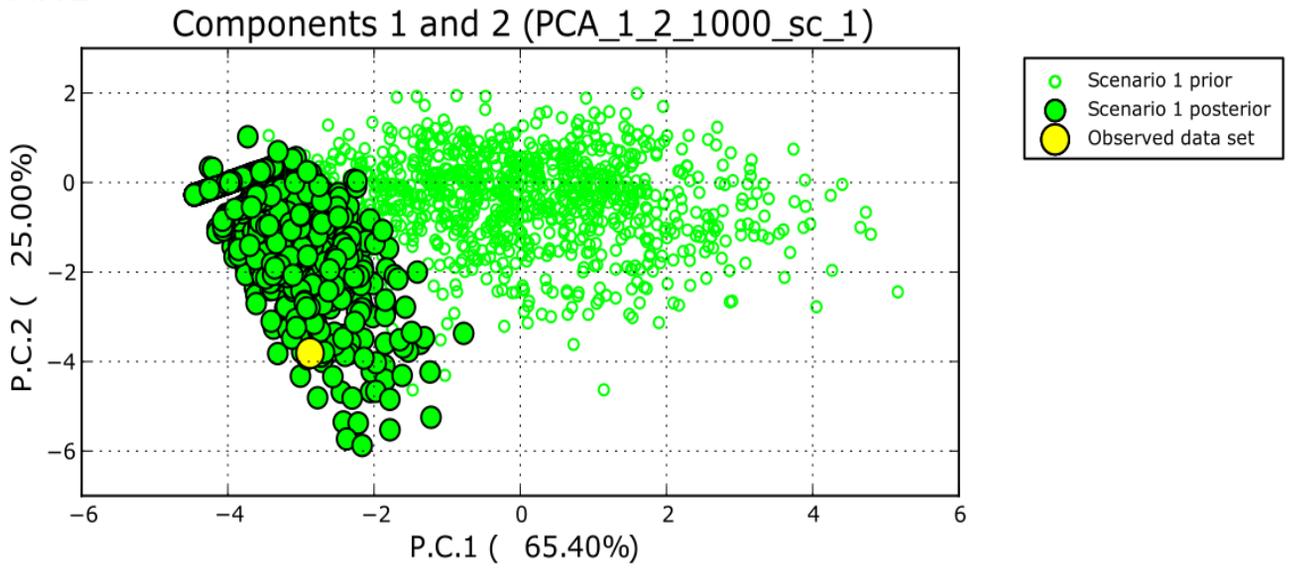
# IZU1



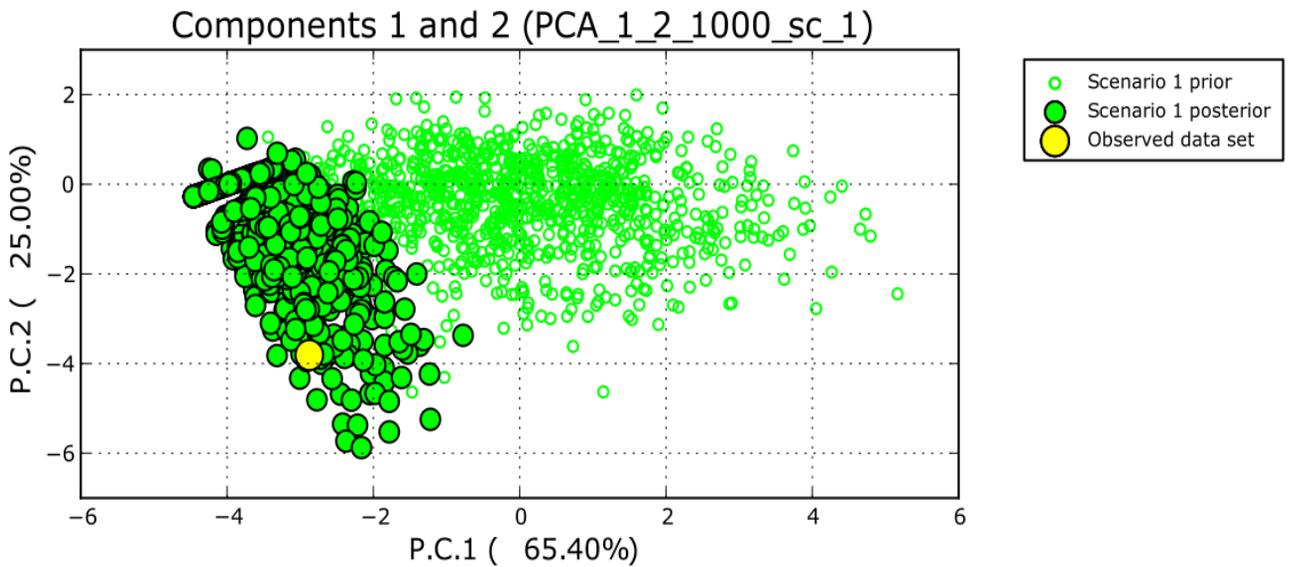




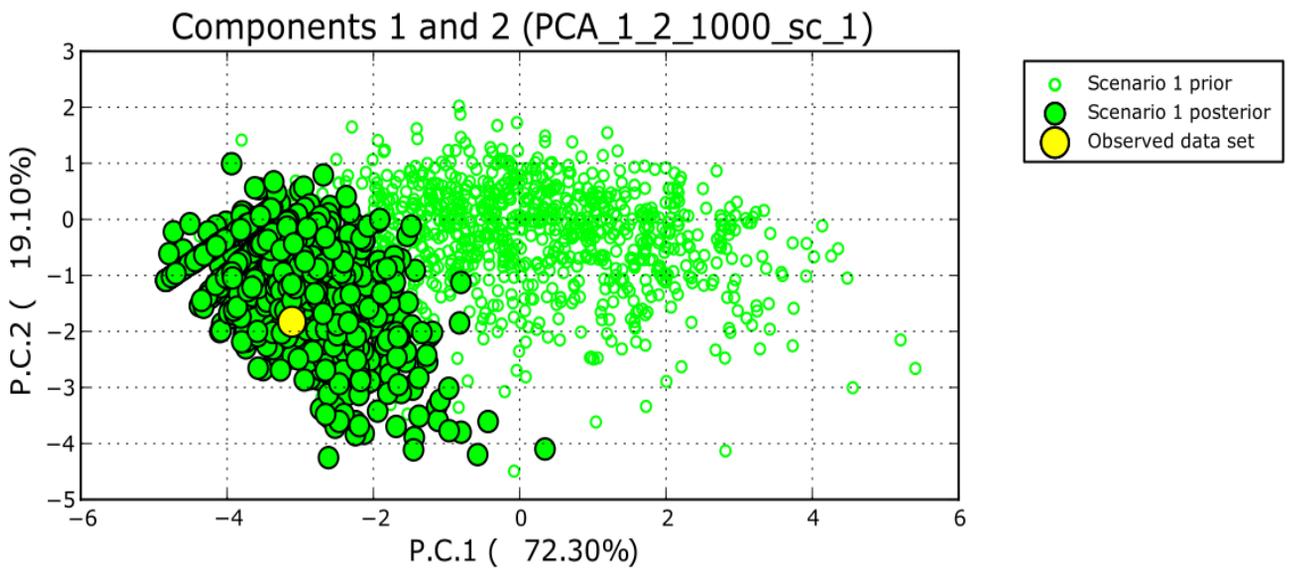
# ESAN1



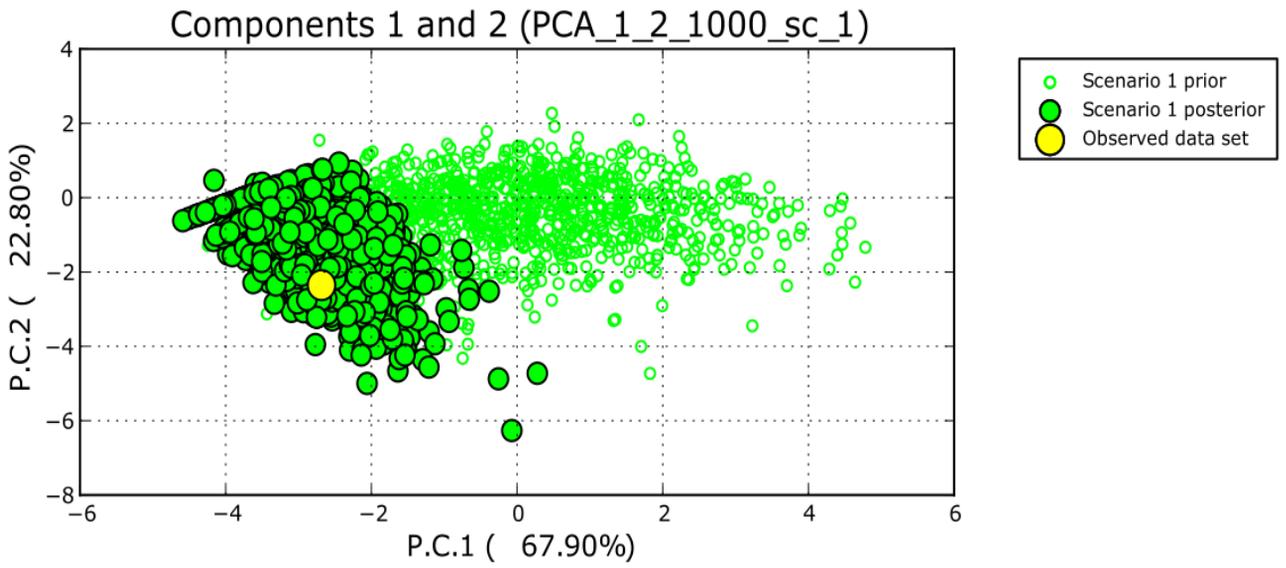
# ESAN2



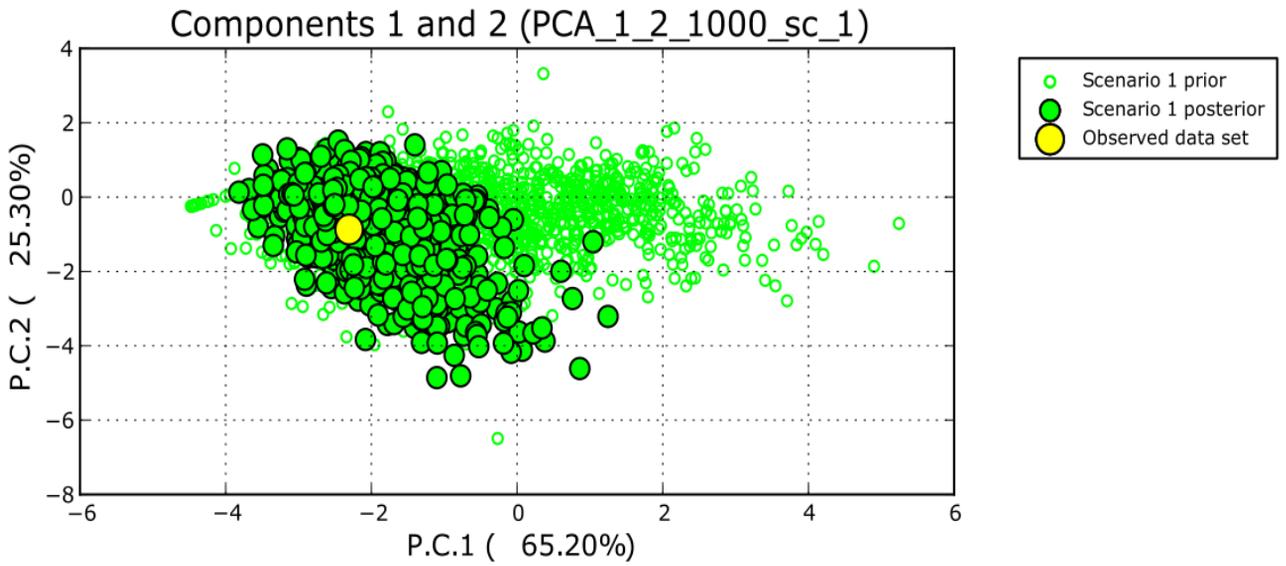
# SAND1



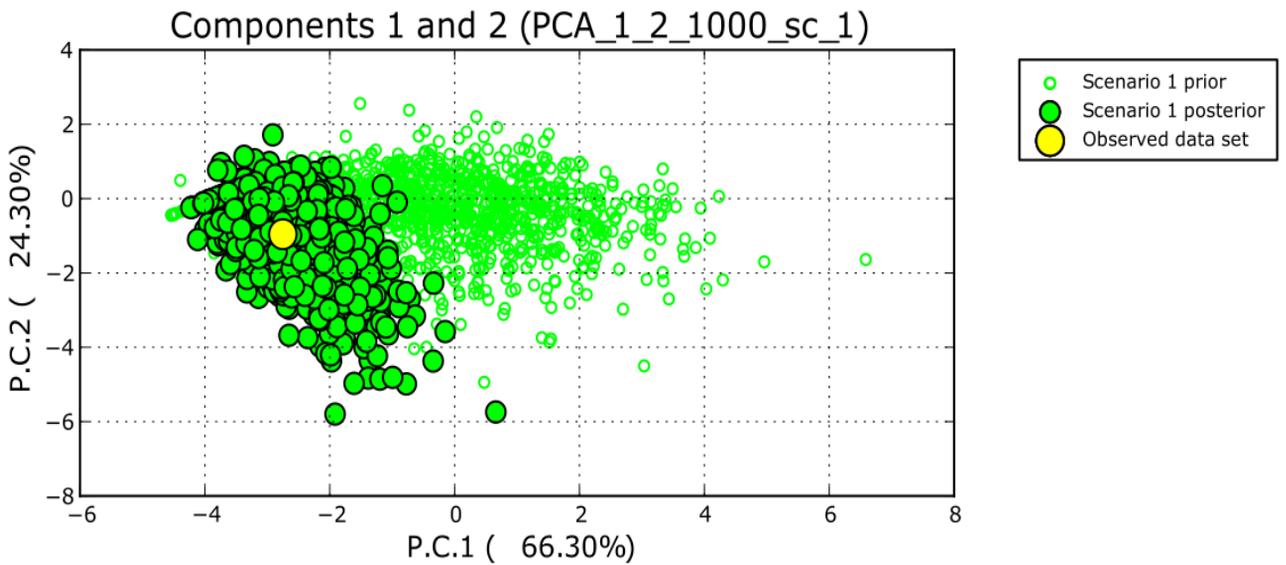
# KANT



# IZU1



# IZU2



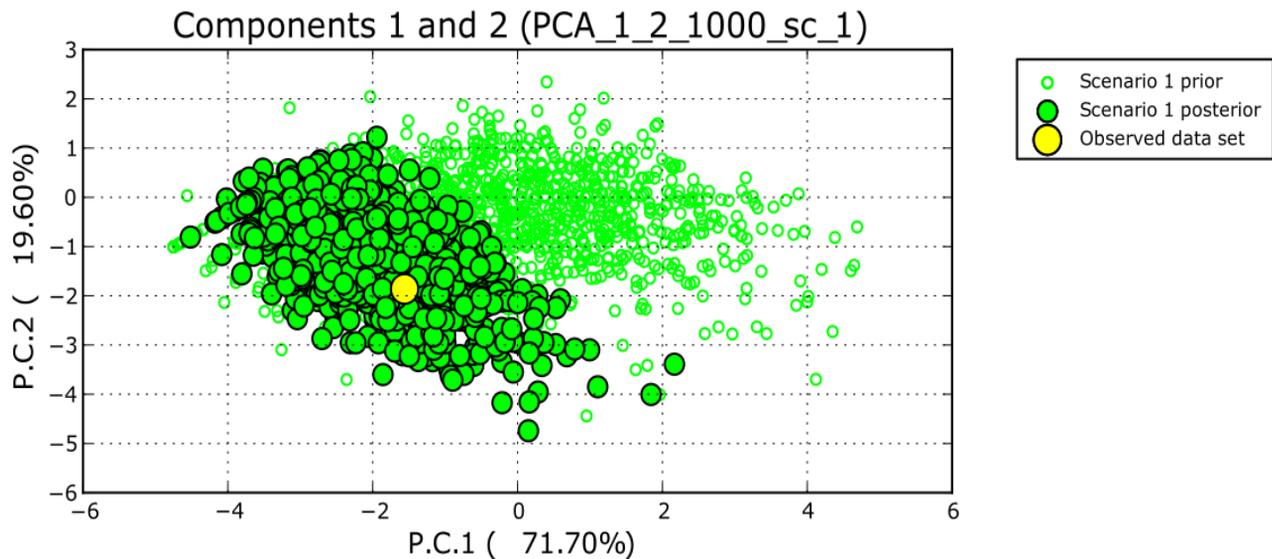


Fig. 1S2

PCA plot of prior, posterior and observed data set

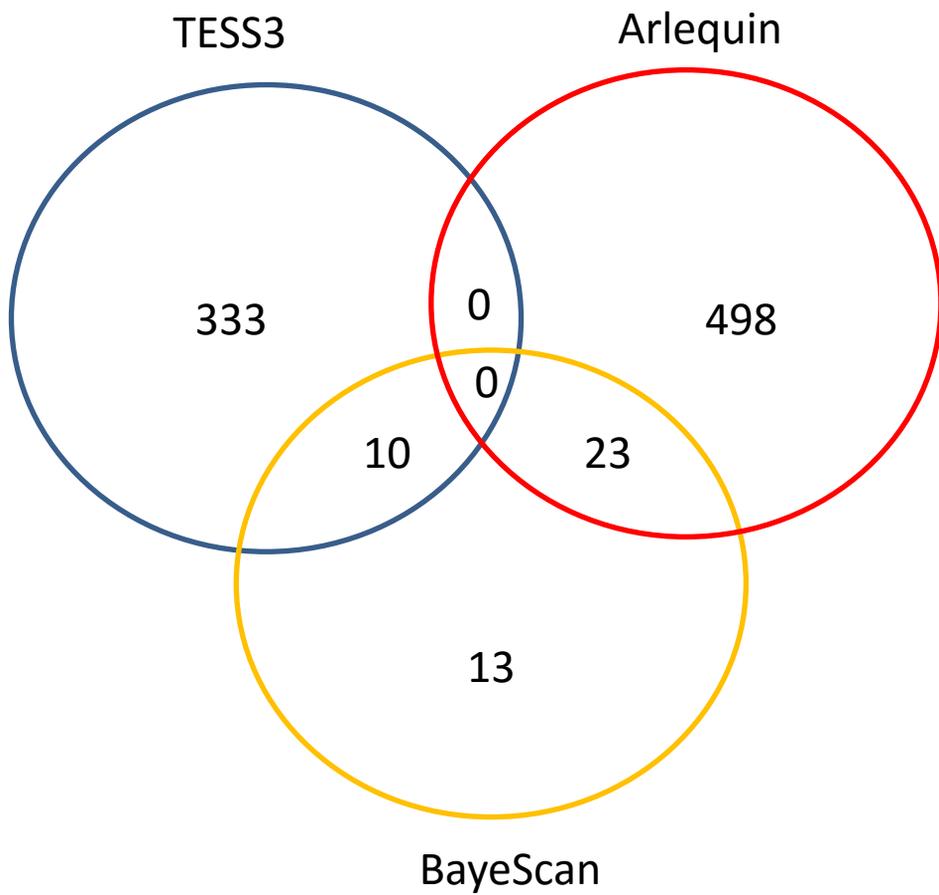


Fig. 2.S1 Number of loci inferred as a outlier loci by three softwares

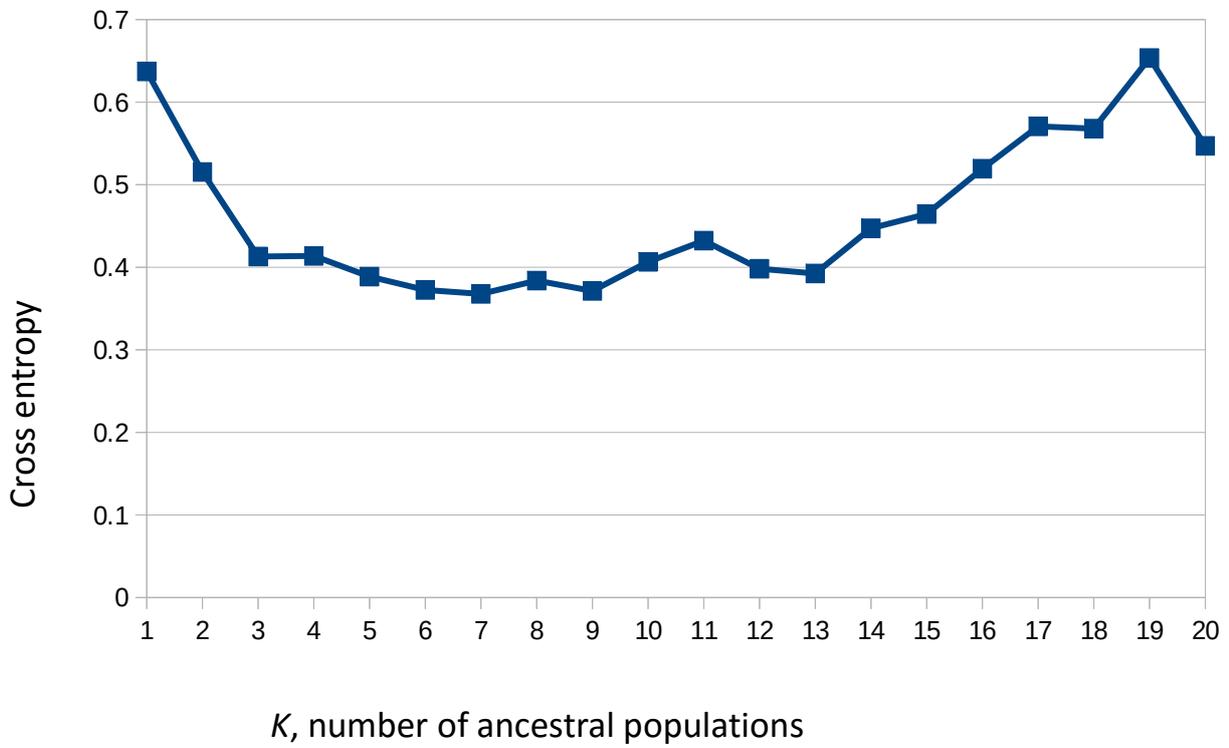


Fig. 2.S2 Number of ancestral populations along the cross entropy. The cross entropy decreases steadily to  $K = 3$ . The cross entropy was smallest in  $K = 7$ .

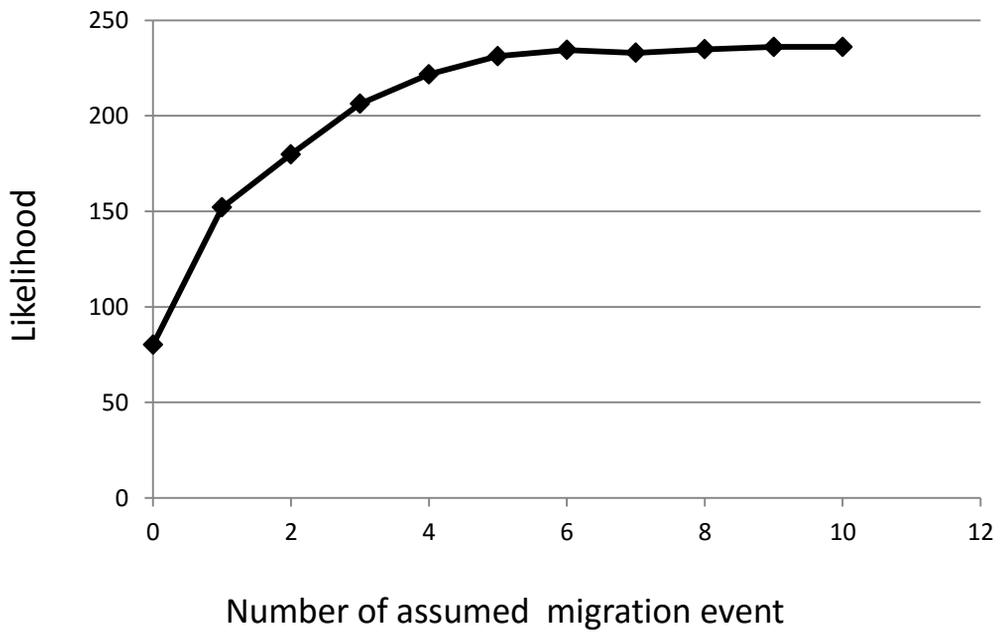
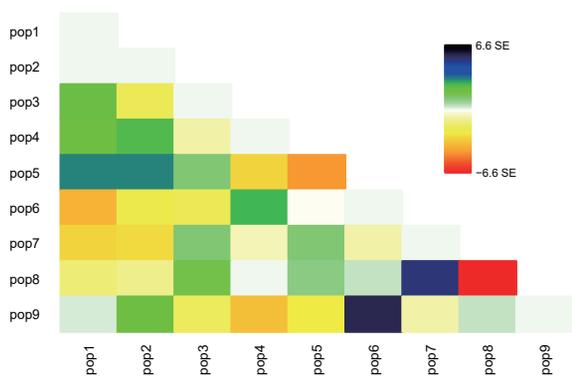
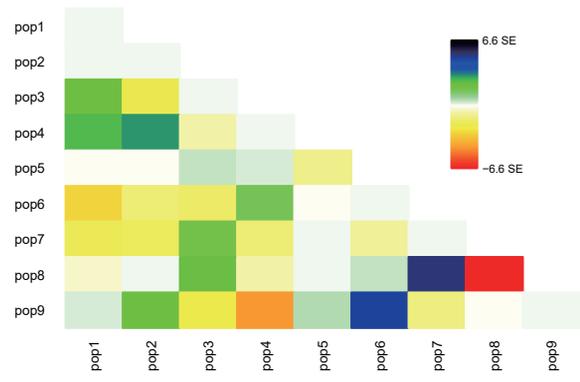


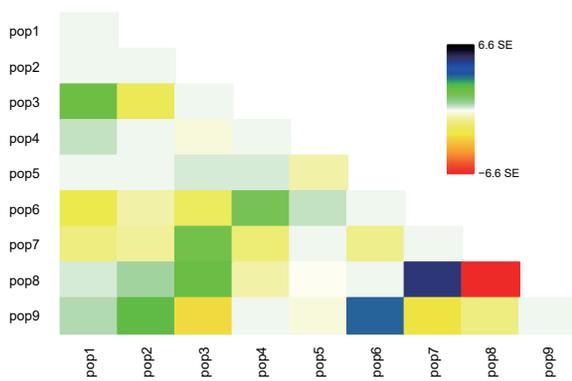
Fig. 2.S3 Likelihood along number of assumed migration event.



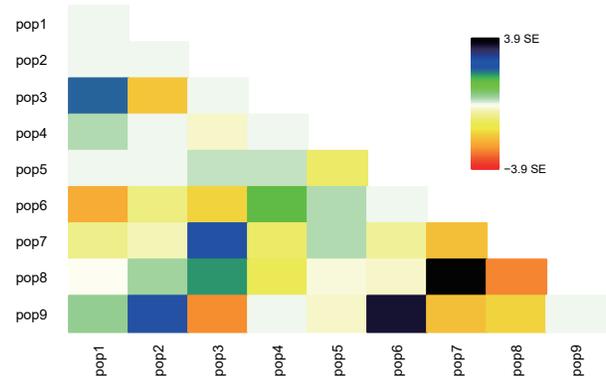
$m = 0$



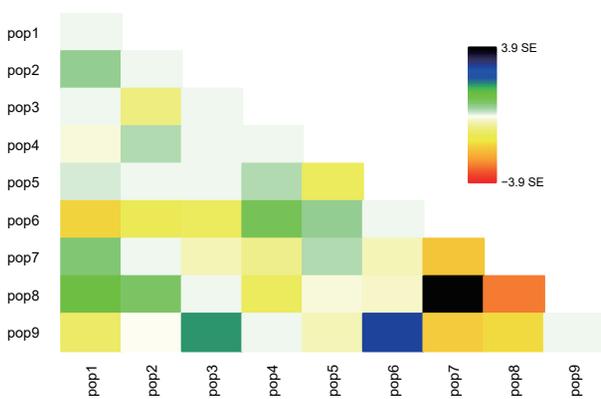
$m = 1$



$m = 2$



$m = 3$



$m = 4$

Fig. 2.S4

Heat map of standard error of drift parameters in each pair of population. Colors indicates excess or deficiency of drift parameters under assumed number of migration events.