doi:10.1111/j.1558-5646.2008.00320.x

THE FREQUENCY OF rDNA VARIANTS WITHIN INDIVIDUALS PROVIDES EVIDENCE OF POPULATION HISTORY AND GENE FLOW ACROSS A GRASSHOPPER HYBRID ZONE

Irene Keller,[1,2,3] Paris Veltsos,[1] and Richard A. Nichols[1]

[1] *School of Biological and Chemical Sciences, Queen Mary, University of London, London, United Kingdom*

[2] *E-mail: keller.irene@gmail.com*

Received August 29, 2007
Accepted December 1, 2007

In the grasshopper *Podisma pedestris*, units of the ribosomal DNA (rDNA) multigene family are not identical, but comprise multiple genetic variants. We surveyed this variation using a novel pyrosequencing approach. The history of the study populations is well characterized as the pattern of colonization can be inferred from the distribution of two chromosomal races that invaded from different directions after the last glacial maximum and finally met to form a hybrid zone. This knowledge of the populations' ancestry allows us to draw inferences about the rate of change in rDNA composition. The rDNA data have, in turn, been revealing about the populations' ancestry, indicating a previously unsuspected route of postglacial colonization. The two chromosomal races were found to have genetically distinctive rDNA composition, demonstrating the persistence of differences for thousands of generations. It follows that the hybrid zone represents a natural experiment in which repeated crossing and backcrossing between these different rDNA lineages has occurred for over 8000 generations. The association between chromosomal race and rDNA composition has been broken down within the zone. It therefore appears that rDNA variants move freely across the zone and are not under opposing selection pressures in the two races, as had previously been suspected.

KEY WORDS: **Concerted evolution, cline, tension zone.**

In eukaryotes, the 45s component of rDNA typically occurs in many hundreds to thousands of copies per genome arranged in one or more tandem arrays. Comparisons between species reveal a pattern known as concerted evolution, in which new mutations arising in one rDNA unit appear to have spread rapidly to all copies in the rDNA family (Elder and Turner 1995; Graur and Li 2000). There has even been speculation that the processes responsible for rDNA concerted evolution could promote rapid divergence between populations within a species and, ultimately, lead to speciation (Liao 1999). In contrast to this widely observed pattern (Eickbush and Eickbush 2007), our previous work on the grasshopper *Podisma pedestris* has shown that its rDNA is not homogenized very rapidly (Keller et al. 2006). In particular, each *Podisma* individual contains many rDNA pseudogenes, some of which have persisted in the genome for millions of years. As well as this intraindividual variation in the coding sequence, differences were also found in the internal transcribed spacer 1 (ITS1) sequence of apparently functional rDNA units.

PYROSEQUENCING AS A TOOL FOR STUDYING MULTIGENE FAMILIES

Given these genetic differences between rDNA units within individuals, a logical next question is whether the relative frequency of variants differs between individuals and populations. This information could potentially provide clues about the origin and

[3] Current Address: Eawag, Department of Aquatic Ecology, Ueberlandstrasse 133, 8600 Duebendorf, Switzerland.

persistence of genetic differences within rDNA. Previous studies of rDNA variation have made use of cloning and sequencing, or restriction digest methods, neither of which is well suited to scoring the frequency of variants within individuals for large sample sizes. Instead, in this article, we introduce the use of pyrosequencing: a quantitative sequencing method that was developed to estimate relative allele frequencies in pools of DNA sequences (Lavebratt et al. 2004). The frequency of each base at sites in which there is a single nucleotide polymorphism (SNP) or a short insertion/deletion polymorphism is inferred from the relative signal intensity of four dyes that correspond to the four bases. In medical microbiology, for example, pyrosequencing has been used to determine the relative frequencies of different genotypes in bacterial isolates (Monstein et al. 2001; Tarnberg et al. 2002; Kolak et al. 2003). In the case of multigene families such as rDNA, the technique can be used on the DNA obtained from a single individual to estimate the relative proportion of different variants among the copies making up the multigene family. The automated allele frequency estimation is many times faster than approaches based on direct sequencing, in which the height of double peaks is compared by eye (e.g., Rauscher et al. 2002).

THE PODISMA POPULATIONS OF THE SOUTHERN FRENCH ALPS

By surveying populations over a wide geographic area, it is possible to obtain insights into the origin and spread of genetic variation. In the southern French Alps, the distribution of two chromosomal *Podisma* races has been interpreted to infer that the area was colonized from at least two different directions after the last ice age, as the climate warmed and the mountains became hospitable for the grasshoppers. Furthermore, there is intriguing preliminary evidence that the rDNA from the two different origins might be incompatible.

Where the two races meet, they form a hybrid zone (Fig. 1A; Barton and Hewitt 1981a). The so-called fused race (named "XY" in Barton and Hewitt 1981a) has the more limited geographic distribution (Fig. 1A) and is distinguished from the unfused race (XO in Barton and Hewitt 1981a) by a Robertsonian fusion involving the X chromosome and an autosome (Hewitt and John 1972; Westerman and Hewitt 1985). Previous studies indicate that the hybrid zone is maintained by a balance of dispersal into the zone and selection against hybrids due to genetic incompatibilities between the races (Barton and Hewitt 1981a,b; Nichols and Hewitt 1986, 1988). However, although breeding experiments have suggested that as many as 150 independent loci may contribute to the reduced viability of hybrids (Barton and Hewitt 1981b), only two consistent racial differences other than the fusion itself have been detected so far. First, the DNA content is considerably higher in the unfused race than in the fused race (Westerman et al. 1987). Second, Dallas et al. (1988) detected a restriction polymorphism in rDNA that was in very strong linkage disequilibrium with the Robertsonian fusion, i.e., the restriction site was present in most individuals with a fused X and absent in most individuals with an unfused X, even in the center of the hybrid zone where fused and unfused chromosomes occur in hybrids and recombine. Silver staining of chromosome spreads showed that, in most individuals, there was an rDNA locus near the telomere of the X chromosome (Gosalvez et al. 1988; Bella et al. 1991). Consequently, an association between the fusion and a particular rDNA sequence variant should be broken up rapidly by recombination unless it is maintained by selection.

The hybrid zone has probably existed at its present location for more than 8000 years since the last climatic optimum (Nichols and Hewitt 1986). After such an extended period of interbreeding, neutral loci will have mixed between the races as a consequence of gene flow (Barton 1979; Barton 1983). Given the dispersal rate of *P. pedestris*, it has been estimated that neutral alleles from one race would, by now, have penetrated several kilometers into the range of the other. Consequently, even if there had been an initial allele frequency difference between the two races, a present-day survey would reveal broad clines (Barton and Hewitt 1981a). Conversely, alleles that were selected against when they entered the range of the other race would show sharp transitions; hence the pattern of allele frequency changes across the hybrid zone can be used to identify the signature of selection.

Here, we investigate these patterns in rDNA by carrying out a detailed survey of the genetic variation at several polymorphic sites (SNPs) in functional rDNA sequences in fused, unfused, and hybrid populations sampled across the Alpes Maritimes to (1) investigate the potential involvement of rDNA in the genetic incompatibility between the two chromosomal races, (2) infer the recolonization history of the area after the last glacial maximum from the observed population structure, and (3) assess the differences in rDNA composition between individuals and populations and relate it to the molecular evolution of these sequences.

Material and Methods

SAMPLES

Samples of *P. pedestris* were collected between 2003 and 2005 in the Alpes Maritimes in southern France. The sampling concentrated on two main regions, one near Seyne at the western end of the hybrid zone and one near Tende close to the French–Italian border (Fig. 1A). We employed a hierarchical sampling design where, in most cases, several populations were sampled both within mountain blocks and within races. Mountains were considered as different blocks if they were separated by valleys consistently below 1500 m above sea level, where the species does not normally occur (Baur et al. 2006; R. A. Nichols, unpubl. data). Only male grasshoppers were sampled as they can be karyotyped

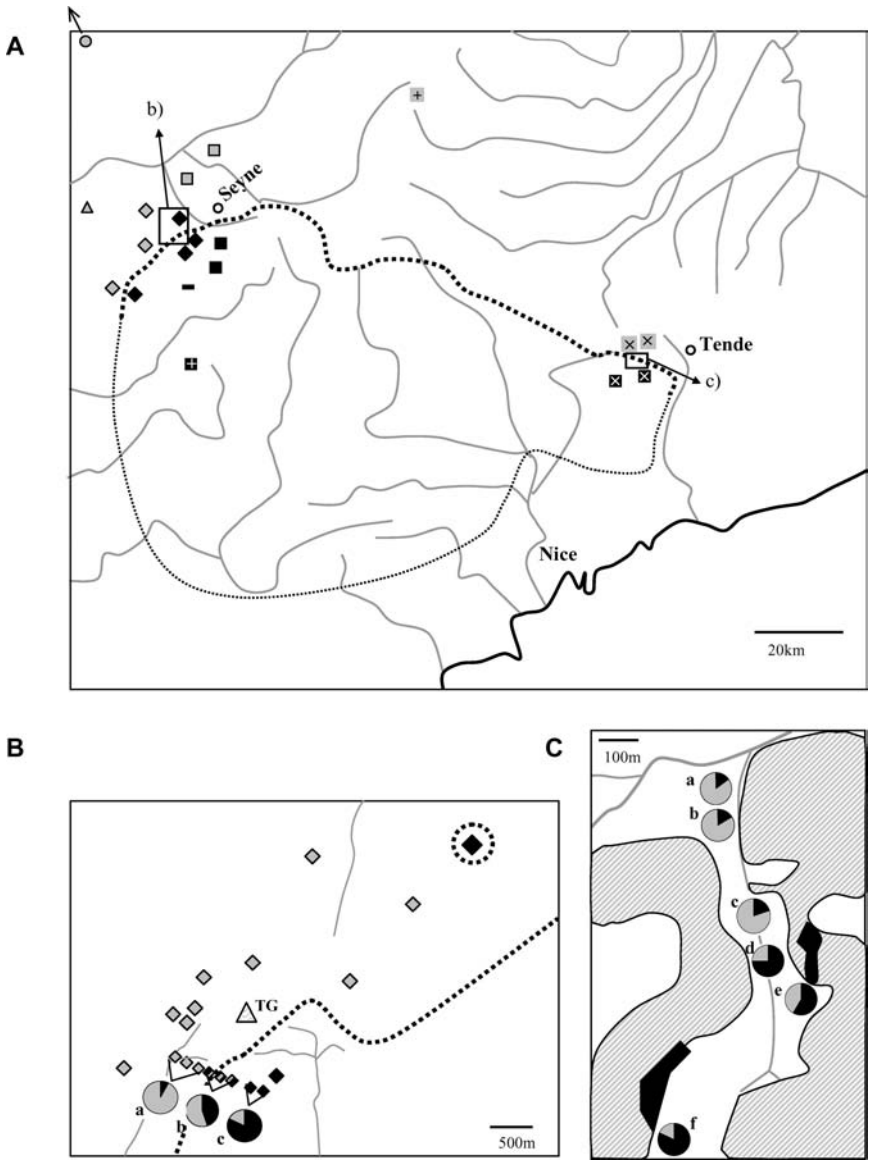

Figure 1. **(A) Sampling locations in the Alpes Maritimes in southern France. The distribution of the fused race is indicated by the dotted line with the hybrid zone in bold (based on fig. 1 from Barton and Hewitt 1981a). Samples from fused populations are shown in black, samples from unfused populations in gray. Different symbols are used for different mountain blocks separated by areas below 1500 m above sea level. The sampling site in the top left-hand corner is near Grenoble outside the area covered by the map. (B) Additional sampling locations in the area around Tête Grosse near Seyne. Samples from predominantly fused populations are shown in black, samples from predominantly unfused populations in gray. Pie charts indicate the frequency of the two karyotypes along a transect across the hybrid zone. Note the isolated fused population north of the hybrid zone. TG, Tête Grosse. (C) Transect across the hybrid zone near Tende. Pie charts show the frequency of fused (black) and unfused (gray) karyotypes at six different sampling locations. Forested areas are hatched and two cliffs are shown in black (map based on fig. 6 from Barton and Hewitt 1981a).**

more easily. The testes were dissected out and fixed in a freshly prepared 3:1 mixture of ethanol and glacial acetic acid. The hind legs were stored in 100% ethanol for DNA analysis. Three to seven individuals per sampling location were used for the population genetic survey (see online Supplementary Appendix S1). In most cases, samples collected at a distance of more than 1 km from the hybrid zone were not karyotyped but assumed to be purely fused or purely unfused, respectively.

In both regions, the exact location of the hybrid zone is known from previous studies (Barton and Hewitt 1981a; Halliday et al. 1984). In 2005, samples were obtained from transects across the zone near Seyne (Fig. 1B) and near Tende (Fig. 1C). Thirteen samples of 10–13 individuals were collected at 100-m intervals across the zone with individuals taken from within a 20-m radius at each point. The karyotypes of all individuals were determined from orcein-stained squash preparations of testicular follicles. For

the population genetic analysis (see below), we randomly selected six individuals of each karyotype from the center of the hybrid zone where the frequency of the fused X was ca. 50% (pie chart b in Fig. 1B and average of pie charts c-e in Fig. 1C). At Tende, we also included four individuals for which the karyotype could not be determined because they were infested with nematode parasites and had poorly developed testes. Two to four individuals of each karyotype were analyzed from both ends of the Seyne and Tende transects. At both locations, the frequency of the fusion was below 25% at the unfused end of the transect (pie chart a in Fig. 1B and average of pie charts a and b in Fig. 1C) and above 75% at the other end (pie chart c in Fig. 1B and pie chart f in Fig. 1C). Each sample included fused and unfused individuals (see online Supplementary Appendix S1). Consequently, some individuals represented the most common karyotype at a particular location and some the rarer karyotype.

In the area west of Seyne, a previous study (Nichols et al. 1990) detected an isolated fused population that was completely surrounded by unfused populations (Fig. 1A,B). To investigate this area in more detail at the sequence level, additional samples were collected around Tête Grosse (Fig. 1B). Sample sizes ranged from two to eight individuals (see online Supplementary Appendix S1 for details) and all individuals were karyotyped as described above.

All sampling sites were located on a Google Earth map (http://earth.google.com) and their coordinates were determined. The program Earth written by J. A. Byers (http://www.wcrl.ars.usda.gov/cec/java/lat-long.htm) was used to calculate the geographic distance between all pairs of sampling sites based on these coordinates. For the smaller distances between the closely spaced populations of the transect, geographic distances were measured directly from Google Earth as this was found to give slightly more accurate results (i.e., closer to the measurements made in the field).

MOLECULAR METHODS

Identification of single nucleotide polymorphisms

Polymorphic sites were identified by sequencing the ITS1 in 58 individuals of *P. pedestris* from six populations of each race and from the Seyne transect. This sequencing effort focused mainly on the western region around Seyne but also included five individuals from Tende (Fig. 1A). The ITS1 was amplified using the primers 18S(f) (5′-CCTTTGTACACACCGCCCGT) and ITS6(r) (5′-GTTCATGTGTCCTGCAGTTCAC) (Sharpe et al. 2000), which generated a 650-bp fragment in *P. pedestris*. This primer pair had been used extensively in a previous study in which we had demonstrated that, in the presence of dimethyl sulfoxide (DMSO), putatively functional rDNA units were the main PCR product (Keller et al. 2006). PCR amplifications were carried out in volumes of 25 μl using a PTC-220 DNA Engine DyadTM Peltier Thermal Cycler (MJ Research, Waltham, MA). Each reaction contained approximately 5–10 ng of genomic DNA, 5% DMSO, 3 mM $MgCl_2$, 0.12 mM of each dNTP, 12.5 pmol of each primer, 0.25 units of BIOTAQ DNA Polymerase (Bioline, London, U.K.), and the associated NH_4 buffer at 1× concentration. An initial denaturation step of 5 min at 95°C was followed by 29 cycles of amplification (30 sec at 94°C, 30 sec at 62°C, 1 min at 72°C) and a final elongation step of 10 min at 72°C. The resulting PCR products were cleaned and sequenced on an ABI 3700 sequencer. The electrophoretograms were then visually inspected for double peaks indicating polymorphic sites. It has been shown that this approach successfully detects variants that occur at a frequency of more than 10% (Rauscher et al. 2002). Pyrosequencing assays were developed for all seven sites that were polymorphic in at least three populations. One of these is a single base pair deletion but, in the following, it will be referred to as SNP 5 for simplicity. SNP 3 has three different alleles whereas all remaining loci are diallelic.

PCR and pyrosequencing

DNA was extracted from muscle tissue of one or both hind legs using a Qiagen DNeasy Tissue Kit (Qiagen, Crawley, U.K.) according to the manufacturer's instructions. A nested PCR approach was used to generate the template for the pyrosequencing analysis. In a first step, ITS1 was amplified as described in the previous section. The use of the extensively tested primers 18S(f) and ITS6(r) guaranteed the preferential amplification of putatively functional rDNA units (Keller et al. 2006). In a second step, a 1:200 dilution of this first PCR reaction was used to generate shorter PCR fragments more suitable for the pyrosequencing assay. The sequence of all primers is given in Table 1. The first pair of PCR primers amplified a 163-bp fragment containing SNPs 1 and 2, the second pair a 177-bp fragment containing SNPs 3 to 5, and the third pair a 162-bp fragment containing SNPs 6 and 7. In each case, the reverse primer was biotinylated to allow the efficient cleanup of the PCR products before pyrosequencing (see below). PCR amplifications were carried out in volumes of 50 μl using a PTC-220 DNA Engine DyadTM Peltier Thermal Cycler (MJ Research). Each reaction contained 3 μl of diluted PCR product from the first step, 1.5 mM $MgCl_2$, 0.2 mM of each dNTP, 10 pmol of each primer, 1.2 units of BIOTAQ DNA Polymerase (Bioline), and the associated NH_4 buffer at 1× concentration. An initial denaturation step of 5 min at 95°C was followed by 45 cycles consisting of 15 sec at 95°C, 30 sec at a primer-specific annealing temperature (Table 1) and 15 sec at 72°C, and a final elongation step of 5 min at 72°C.

Pyrosequencing primers for all seven SNPs were designed by Biotage and separate assays were run for each locus. For each assay, 40 μl biotinylated PCR products were combined with 3 μl streptavidin sepharose beads (GE Healthcare, Amersham, U.K.)

Table 1. PCR and pyrosequencing primers used for SNP genotyping. T_A, annealing temperature used in the PCR.

Primer name	Primer type	Primer sequence 5′–3′	T_A
PCR1-forw	PCR	GTAGGTGAACCTGCGGAAGGA	56°C
PCR1-rev	PCR	*bio*-TGGTGCGTTAGTTTTGAGAGG	
PyroSNP1	Pyrosequencing	GCGGAAGGATCATTACC	
PyroSNP2	Pyrosequencing	CATTTGACACCACCTGTA	
PCR2-forw	PCR	ACGCGCTCCTCTCAAAACTA	56°C
PCR2-rev	PCR	*bio*-CTCGTCGTACGGCTCCAC	
PyroSNP3	Pyrosequencing	CTAACGCACCAGAGTTT	
PyroSNP4	Pyrosequencing	CAGTGTACGGAGGTGG	
PyroSNP5	Pyrosequencing	ATGACGTCCGCCCTA	
PCR3-forw	PCR	TGGCACACACTGGAACAGG	58°C
PCR3-rev	PCR	*bio*-ACATGTCGACGCGCAATT	
PyroSNP6	Pyrosequencing	GCCCCGGTCCCGCCT	
PyroSNP7	Pyrosequencing	CAAACGAAACAATGAAA	

in 37 μl binding buffer (Biotage, Uppsala, Sweden) and incubated on a shaker for 10 min. The beads were then aspirated with a vacuum prep tool (Biotage) and washed for 5 sec in 70% ethanol, 5 sec in NaOH (0.2M), and 10 sec in wash buffer (Biotage). Then, they were released into a PSQ 96 Plate Low (Biotage) containing 40 μl annealing buffer (Biotage) and 5 μl pyrosequencing primer at a concentration of 3 mM. The plate was incubated at 80°C for 2 min and analyzed on a PSQ96MA machine using Pyro Gold reagents (Biotage) as specified by the manufacturer. The dispension order of the four nucleotides was determined automatically by the software. Allele quantification was also performed automatically using the default settings.

A pilot study confirmed that the pyrosequencing approach provided repeatable results (SD between replicates less than 3% in all cases), which were, on average, within 10% of the estimates obtained from peak heights in sequencing electrophoretograms. Specifically, we compared the standardized signal intensity of a particular nucleotide in a double peak with the standardized signal intensity observed at the same position in a monomorphic sequence (i.e., a cloned sequence from the same individual). For standardization, the signal intensity of a particular peak was divided by the mean signal intensity of 10 single peaks of the same nucleotide upstream of that position. In total, we analyzed 14 polymorphic sites in three different individuals. Generally, it is difficult to determine if a low pyrosequencing peak is due to a low frequency allele or background noise. Consequently, when discussing patterns of polymorphism, allele frequencies below 5% are considered artifactual.

STATISTICAL ANALYSES

Principal component analysis (PCA) was performed on the allele frequency data to investigate spatial patterns. One allele per SNP was omitted from the analysis and the frequency estimates from each individual were arcsine transformed for normalization (c.f. Sokal and Rohlf 1995). The PCA was then carried out on the correlation matrix in R (R Development Core Team 2007). For clarity, separate plots of the first two principal components (calculated from the full dataset) were created for subsets of the samples. First, the genetic structure across the entire study area was investigated using a widely spaced set of samples mapped in Fig. 1A (level 1). Then, the area around Tête Grosse west of Seyne was studied in more detail based on the small-scale sampling mapped in Figure 1B and all additional samples from this particular mountain block, which are indicated by gray and black diamonds in Figure 1A (level 2). Finally, analyses at a very fine geographic scale included data from the two transects only (levels 3a and 3b).

To test for isolation by distance, pairwise Mahalanobis distances were computed from the rDNA data for each of the three geographic scales discussed above. The Mahalanobis distance is particularly suitable for this dataset as it does not assume independence of loci. Note that the inference of haplotypes is not possible using pyrosequencing and would require the cloning and sequencing of a large number of rDNA units from each individual. Both the matrix of pairwise Mahalanobis distances and the matrix containing pairwise geographic distances were log transformed. A geographic distance of 1 m was assumed for individuals collected at the same location. A Mantel test was performed in R based on the Pearson correlation coefficient with 10,000 permutations. To investigate the effect of race or mountain block after accounting for geographic distance a partial Mantel test was carried out using the same settings.

Visual inspection of a plot of geographic versus genetic distance for all samples included at level 1 suggested that fused individuals tended to be genetically more similar than unfused individuals or pairs from both races. To investigate this statistically, the following permutation test was developed in R. A linear regression was carried out between pairwise geographic distances

and pairwise Mahalanobis distances. All pairwise comparisons were then assigned to one of three categories based on whether they involved two fused individuals, two unfused individuals, or one individual from each race and the mean of the residuals was calculated for each category. These categories were then randomized and the mean of the residuals was recomputed. This procedure was repeated 1000 times and the observed means were compared to the 2.5% and 97.5% quantiles of the simulated null distribution.

Halliday et al. (1983) had carried out a detailed population genetic study of the Tende area based on 12 allozyme loci. We reanalyzed their data (tables 2 and 3 in Halliday et al. 1983) by carrying out a PCA using one allele per locus. The influence of geographic distance and race on genetic distance was investigated using Mantel tests and partial Mantel tests as outlined above. Following Halliday et al. (1983), we used Nei's standard genetic distance that was computed in GenDist from the Phylip package (http://evolution.genetics.washington.edu/phylip.html). The geographic distance between all pairs of sampling locations was inferred from figure 1 in Halliday et al. (1983).

Results

PATTERNS OF GENETIC VARIATION

For this study, we had selected SNPs that were polymorphic in a set of populations drawn from throughout the Alpes Maritimes, equally divided between fused and unfused sides of the zone (see Materials and Methods). The full set of populations shown in Figure 1 was then surveyed. At most SNPs, one variant was very common and an alternative nucleotide reached medium-to-high frequencies in some individuals in some populations. There were no fixed differences between the two races but some variants were clearly more common in one race than the other.

The frequency of the SNP variants differed between the individuals within any one population, the greatest differences being observed when the average frequencies were around 0.5 (SD between individuals within populations ranged up to 0.4). Some differences between individuals were even more marked. The most striking case was individual 04–9-1 from a population near the Tête Grosse transect (Fig. 1B), in which SNPs 1 and 3 were fixed for a variant that was otherwise not detected in other individuals from this population (see online Supplementary Appendix S1). Similar cases in which individual genotypes differed markedly from the population average were also detected in some other populations (e.g., individuals 04–5-7, 04–12-1, 05–20-1, two individuals each in populations 05–12 and 05–14 in the online Supplementary Appendix S1).

POPULATION STRUCTURE

Level 1: Broad geographic scale

Genetic differentiation depends on geographic distance and chromosomal race. The principal component analysis (PCA) revealed a clear phylogeographic signal with individuals from a geographic area grouping together (Fig. 2). The

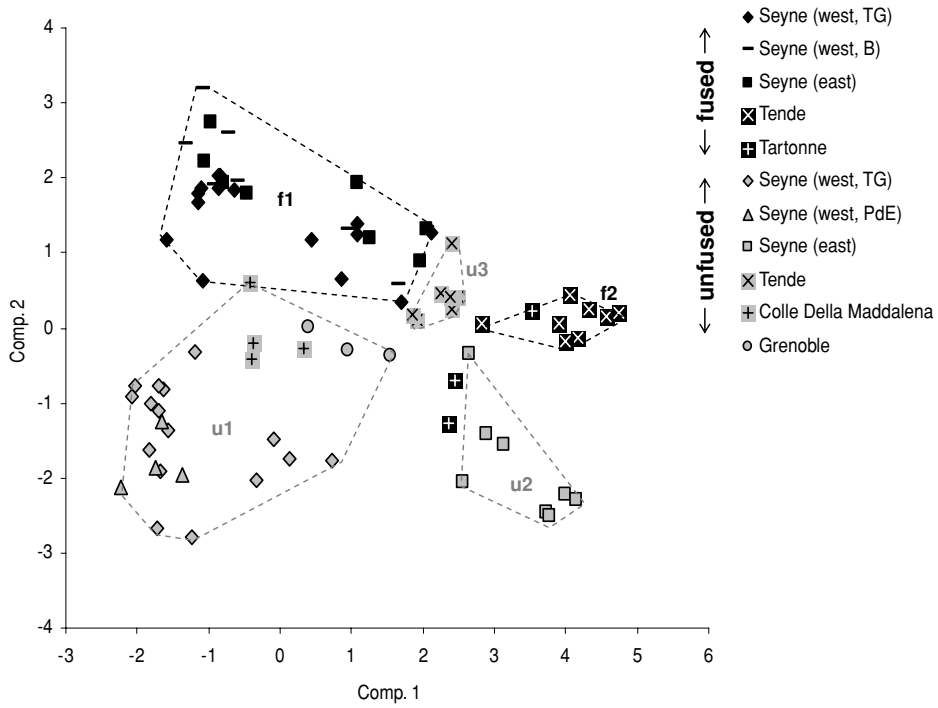

Figure 2. **Level 1: Plot of the first two principal components. Only samples from locations away from the zone (Fig. 1A) are shown. Each point represents an individual and the symbols are analogous to Fig. 1A. Nonoverlapping polygons are drawn around samples from the same race. TG, Tête Grosse, B, Blayeul, PdE, Pente d'Eyrolle.**

samples from the fused race formed two main clusters representing the two sampling areas of Seyne (cluster f1; Fig. 2) and Tende (f2), respectively. Individuals from the Tartonne population appear to be misplaced with unfused populations on this two-dimensional plot, but are in fact displaced on higher axes of the PCA (mean Mahalanobis distance from the nearest cluster center exceeds the mean intracluster value).

In the unfused race, the samples from Tende grouped together (u3; Fig. 2) whereas, unexpectedly, those from Seyne were divided into two distinct clusters (u1 and u2). Cluster u2 contained the samples from the two locations east of the Seyne valley (indicated by squares in Fig. 1A); cluster u1 contained all samples from the mountain blocks west of the Seyne valley (indicated by diamonds and a triangle in Fig. 1A)—plus more distant samples from Grenoble and Colle della Maddalena.

We observed a significant positive relationship between pairwise geographic and genetic distances (Mantel test, $r = 0.31$, $P < 0.001$), although mountain block was not a significant factor after accounting for the effects of geographic distance (partial Mantel test, $r = 0.007$, $P = 0.41$). Individuals belonging to different races, on the other hand, tended to be genetically more different than individuals from the same race (partial Mantel test, $r = 0.20$, $P < 0.001$). The relationship between geographic and genetic distance is illustrated in Figure 3 for pairwise comparisons within each race and between the races, respectively. The increase of genetic distance with geographic distance is linear for the first few kilometers but the patterns become very irregular at a larger spatial scale (Fig. 3).

The fused race is genetically more homogenous than the unfused race. Separate analyses reveal significant patterns of isolation by distance within each race (Mantel test, fused only: $r = 0.17$, $P = 0.001$, unfused only: $r = 0.36$, $P < 0.001$). Graphical representation of these patterns suggested that the fused race was genetically more homogenous than the unfused race (Fig. 3A and 3B). The statistical significance of this difference between the races was confirmed in a permutation test based on the residuals from a linear regression between geographic and genetic distance, which showed that fused individuals were, on average, genetically more similar than predicted from the geographic distance between them. The observed mean of the residuals of −0.221 was considerably lower than the 2.5% quantile of the simulated null distribution (−0.078). The opposite trend was detected for pairs of unfused individuals and between-race comparisons where the observed mean of the residuals was larger than the 97.5% quantile of the simulated null distribution (unfused: mean = 0.083, 97.5% quantile = 0.072; between races: mean = 0.064, 97.5% quantile = 0.009).

Level 2: Local sampling around Tête Grosse

Gene flow across the hybrid zone. A PCA of all samples from the mountain block around Tête Grosse to the west of the Seyne valley showed that individuals collected close to or in the hybrid zone fell between the two distinct clusters formed by individuals collected at least 1 km away from the center of the zone (results not shown). This pattern illustrates that the genotypes of such hybrid individuals were intermediate compared to the parental races. The genotypes of the individuals from the isolated fused population (Fig. 1A,B) were typical of fused individuals from this area.

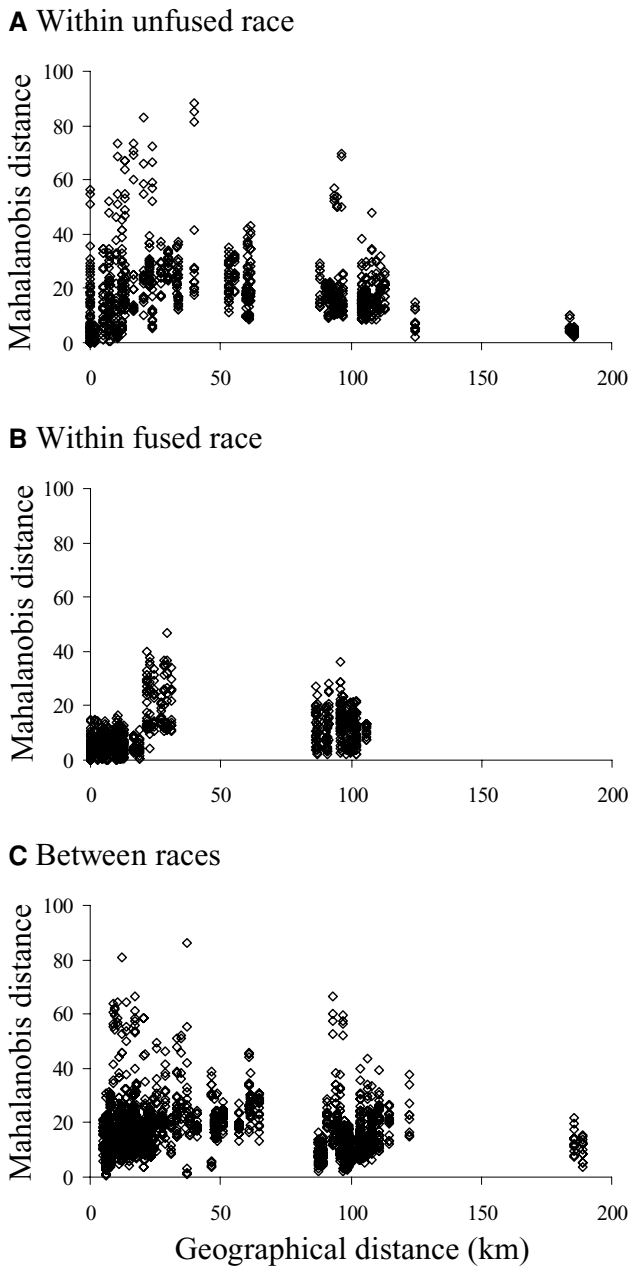

Figure 3. Level 1: Plots of geographic versus genetic distance for pairwise comparisons within the unfused race (A), within the fused race (B) and between the races (C). Only samples from locations away from the zone (Fig. 1A) are included.

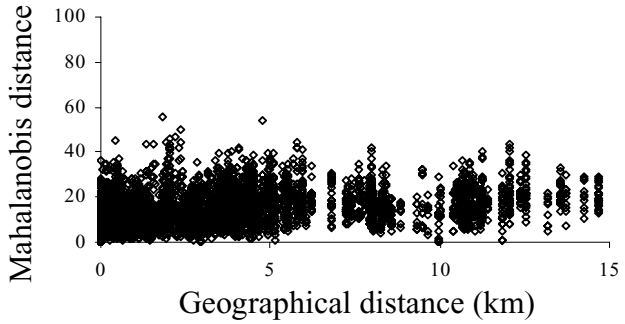

Figure 4. Level 2: Plot of geographic versus genetic distance for pairwise comparisons of individuals from the same mountain block near Seyne. One individual (04–9-1; see online Supplementary Appendix S1) with a very unusual genotype was omitted.

At this geographic scale, we detected a highly significant pattern of isolation by distance (Mantel test, $r = 0.20$, $P < 0.001$). The increase of genetic distance with geographic distance was almost perfectly linear (Fig. 4) without the erratic variation observed at a larger spatial scale (Fig. 3). The effect of race was relatively weak but statistically significant, with individuals from the same race being genetically more similar, on average, than individuals from different races (partial Mantel test, $r = 0.07$, $P = 0.013$).

Level 3: Transects across the hybrid zone

No association between sequence variants and karyotype. If different variants were favored in the two races we would expect to observe some association between rDNA variant and karyotype in individuals sampled from transects across the hybrid zone. There was no such association in plots of the first two principal components: the hybrids did not group according to karyotype (Fig. 5). There was also no obvious pattern if the individuals were distinguished based on where they had been collected in the zone (i.e., fused side, center or unfused side). In agreement with this finding, neither geographic distance nor race explained a significant proportion of the variance in pairwise genetic distance (Mantel test and partial Mantel test, $P > 0.05$ in all cases).

The relative frequency of some variants differed between pure fused and unfused populations, at SNP 4, 6, and 7 in the Seyne region and 3 and 5 in Tende (see online Supplementary Appendix S1). This is also illustrated by the loadings of the first two principal components (Table 2). SNPs 3 and 5 strongly influence component 1 that differentiates between the two races at Tende, whereas the remaining SNPs mainly influence component 2 that differentiates between the races at Seyne (Fig. 2). The frequencies of different variants did not change rapidly at the chromosomal hybrid zone. For example, at Tende, the allele frequency differences between the two sides of the zone were particularly pronounced

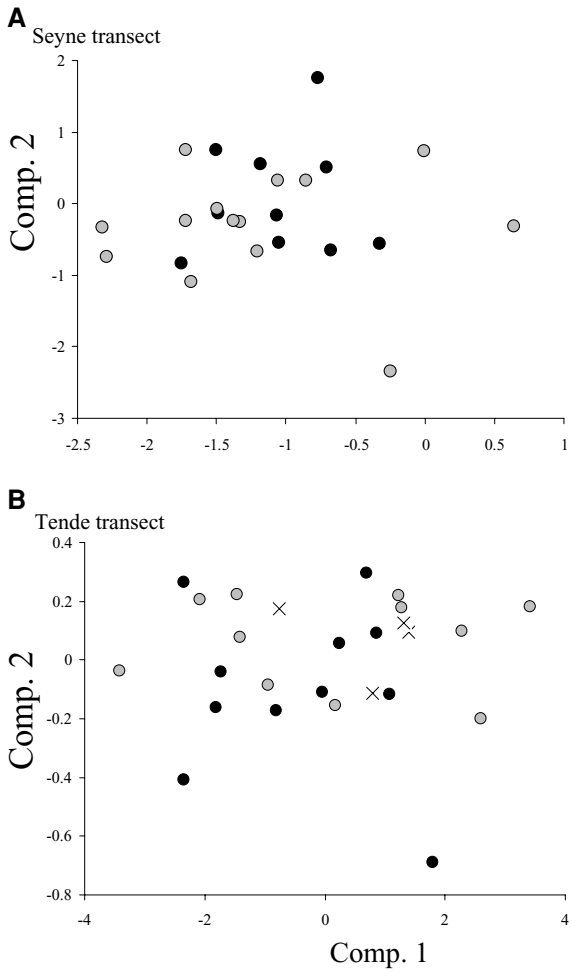

Figure 5. Level 3: Plot of the first two principal components from PCAs based on two transects across the hybrid zone at Seyne (A) and Tende (B). Black circles, hybrids with a fused X chromosome; gray circles, hybrids with an unfused X chromosome; crosses, parasitized individuals with unknown karyotype.

(populations 05-22 to 05-25, see online Supplementary Appendix S1), yet remained quite constant across the narrow (800 m wide) chromosomal hybrid zone.

Tende: allozyme data

Allozymes show the same pattern of genetic variation as rDNA. Because there is only one copy of each allozyme locus per haploid genome, allele frequencies cannot be estimated within individuals but only within populations. A plot of the first two principal components from a PCA showed that the populations tended to group according to race although there was some overlap (not shown). In agreement with the results of Halliday et al. (1983), we detected a significant association between geographic and genetic distance (Mantel test, $r = 0.25$, $P = 0.012$). Our reanalysis has established that the genetic distance between

Table 2. Loadings of the first two principal components.

	Comp. 1	Comp. 2
SNP 1	−0.196	0.393
SNP 2		0.371
SNP 3a	−0.505	0.186
SNP 3b	0.400	0.155
SNP 4	0.353	0.393
SNP 5	−0.518	0.125
SNP 6	0.336	0.401
SNP 7	−0.198	0.564

populations of the same race was smaller, on average, than between populations belonging to different races after allowing for the geographic distance between them (partial Mantel test, $r = 0.28$, $P = 0.001$).

Discussion

In multicopy families like rDNA, variation can be observed at three different levels: within individuals, within populations, and between populations. Previous theoretical (Ohta and Dover 1983) and empirical work has focused on the rDNA diversity and evolution within individuals. Where individuals have been explicitly sampled from more than one population within a species, the data have not been sufficient to establish the extent of population differences (e.g., Parkin and Butlin 2004; Keller et al. 2006). Other studies have made between-population comparisons, but have not assessed the frequency of different rDNA variants within individuals in detail (Crease 1995; Nieto Feliner et al. 2004). The downplaying of intraindividual variation extends to most phylogenetic studies, in which it is assumed to be negligible (see e.g., Álvarez and Wendel 2003). Here, we have used a novel pyrosequencing approach to investigate the patterns of rDNA sequence variation at all three levels: within individuals, within populations, and between populations.

To the best of our knowledge, theoretical studies have not investigated the expected pattern of variation in a subdivided population let alone a subdivided population with a realistic history of colonization. However, it seems likely that sufficiently rapid homogenization of rDNA units would erode historical patterns of variation. At slower rates of homogenization, like those in *Podisma*, the observed rDNA variation may more closely reflect the history and present-day structure of populations. For example, nearby populations may be genetically more similar than more distant ones producing a pattern of isolation by distance.

THE RATE OF rDNA EVOLUTION

The pyrosequencing assay provided repeatable estimates of the relative frequencies of rDNA variants in the genome. Within populations, we detected between-individual variation in the relative frequencies of different variants (see online Supplementary Appendix S1), which was many times larger than the measurement error (SD between replicates = 3%).

On average, the genotypes of individuals from the same population were similar but there is also some evidence that changes may occasionally occur in bursts. There were some cases in which the frequencies of particular variants within an individual differed remarkably from the population mean. The most striking example was individual 04–9-1 (see online Supplementary Appendix S1), where the frequencies of variants at four of the seven SNPs were widely different (by as much as 100% in some cases) from the values observed in all other seven individuals from this location. This result was confirmed by repeating all PCR reactions and the pyrosequencing analysis and cannot be explained by a confusion of samples because a similar genotype was not observed in any other population. This phenomenon appears to be part of a wider pattern: unusual genotypes at one or several SNPs were observed in at least seven other anomalous individuals.

Such differences between individuals could be the result of the presence or absence of entire rDNA loci. Indeed, fluorescent in situ hybridization (FISH) experiments revealed large differences in number and chromosomal location of rDNA arrays often within the same population (P. Veltsos, unpubl. data). However, individuals with an unusual genotype did not necessarily have unusual rDNA loci and vice versa. The genotypes of individuals 05–14-1 and -2, for example, differed considerably from those of 05–14-3 and -4 at SNPs 3, 5, and 7 (see online Supplementary Appendix S1). The number and location of rDNA arrays, however, was identical in all four individuals (one X locus and one telomeric locus; P. Veltsos, unpubl. data).

Further evidence comes from comparison between the two races. The variance of the SNP frequencies within individuals from the same population was, on average, three times larger in the fused (0.023) than in the unfused race (0.007) even though the number of rDNA loci was much more consistent in the fused race (fused: 3 to 5 loci; unfused: 5 to 13 loci). These observations again suggest that the association between number and location of rDNA loci and SNP allele frequencies cannot be very strong.

Evidence of rapid change in rDNA sequences has also been reported in polyploid plants. For example, Kovarik et al. (2004) describe the complete loss of one of the two parental variants introduced into a synthetic tobacco allopolyploid line. It is possible that a sporadic but rapid molecular process, not yet fully characterized, generated these dramatically different genotypes. In *Podisma,* the explanation cannot involve wholesale replacement of rDNA arrays, because polymorphisms in functional units have persisted for extended periods of time (estimated to be at least 3.7 million years) in its genome (Keller et al. 2006).

RECOLONIZATION FROM MULTIPLE REFUGIA AFTER LAST GLACIAL MAXIMUM

The two chromosomal races of *Podisma* probably diverged in allopatry, during a period when the species range was displaced from its current distribution by past climate change. The races will have repeatedly been isolated and then come into secondary contact during the cycles of climate change, and the most recent contact will have occurred when their ranges expanded after the last glacial maximum (c 10,000 years BP). The extended periods of isolation between the two races are likely to have lead to genetic divergence at many loci throughout the genome including the rDNA. This interpretation is supported by the genetic distances between the races, which are larger than would be expected from the geographic separation of the samples (partial Mantel test, $P < 0.001$). The same pattern was observed in our reanalysis of the allozyme data from Tende (partial Mantel test, $P = 0.001$).

The recolonization of the area after the last glacial maximum was probably completed ca. 8000 years ago (Nichols and Hewitt 1986) and populations on different mountain blocks will have been isolated ever since, given the poor dispersal ability of the wingless *P. pedestris* (Barton and Hewitt 1982) and its absence from areas below ca. 1500m (Baur et al. 2006; R. A. Nichols, unpubl. data). However, the rDNA genetic distances between grasshoppers on different mountain blocks were not significantly greater than those between comparable individuals on a single block (partial Mantel test, $P = 0.41$). This suggests that different mountain blocks have been recolonized from the same source population and that the rate of genetic divergence is too slow for the signature of common ancestry to be eroded even in the absence of on-going gene flow.

The history of the fusion is thought to trace back to a single small population in which the rearrangement became established by genetic drift before spreading over a larger area, as the species range expanded. Under this scenario, the fused populations would be genetically similar even in the absence of ongoing gene flow. The rDNA results indicate that the fused race is indeed genetically more homogenous than the unfused race (permutation test). However, the PCA did reveal some substructure, with two distinct clusters of genotypes corresponding to the Seyne and Tende sampling areas (f1 and f2; Fig. 2).

Such a pattern can be explained if the range of the fused race expanded and contracted multiple times. The Tende and Seyne areas could then have been recolonized from genetically isolated sources (e.g., Italy and southern France) producing the observed pattern of genetic differentiation. A similar scenario has been postulated for the grasshopper *Chorthippus parallelus*. For this species, it seems likely that Italy was colonized from the Iberian peninsula during an interglacial but that the Italian and Iberian populations became isolated again in later cold periods (Flanagan et al. 1999).

The pyrosequencing results also revealed unanticipated levels of substructuring within the unfused race and suggest that the Alpes Maritimes may have been recolonized from three distinct unfused ancestral populations after the last glacial maximum. One population may have expanded from the west colonizing the mountain blocks of Pente d' Eyrolle and Tête Grosse (u1; Fig. 2). The expansion of this western ancestral population seems to have been halted by the Vallée de la Blanche, which runs from north to south and is consistently below 1350 m. On the eastern side of this valley, different genotypes were observed which formed the u2 cluster in the PCA (Fig. 2). The samples from Tende formed a third cluster (u3; Fig. 2) suggesting that the eastern Alpes Maritimes were colonized from yet another unfused source population.

This subdivision is also supported by the genomic location of rDNA revealed by FISH. At Seyne, individuals from the unfused eastern populations always have rDNA loci on the X chromosome and the telomere of an autosome, both of which are absent in the unfused western populations. Consistent differences in the number and chromosomal location of rDNA loci were also detected between Seyne and Tende in both races (P. Veltsos, unpubl. data).

EVIDENCE OF GENE FLOW ACROSS THE HYBRID ZONE

In all three types of pairwise comparisons (i.e., within unfused, within fused and between races), the increase in rDNA genetic distance with geographic distance was roughly linear for the first few kilometers, but showed no consistent pattern at longer distances (Fig. 3A, 3B, and 3C) most likely because of the influence of founder events and distinct histories of colonization.

The trend over shorter distances can be explained by gene flow. In fact, the hybrid zone is particularly suitable for demonstrating such a signature of gene flow, because it appears to have formed as a result of secondary contact between two genetically diverged populations (Barton and Hewitt 1981a, 1985) and differences in the relative frequency of different rDNA variants have been detected between the two sides (previous section). However, if such differences between the races are not maintained by selection they will be gradually eroded by gene flow. Barton and Hewitt (1981a) estimated that a neutral frequency cline would now be at least 3.7 km wide if the two races came into contact 5000 generations ago. This implies that an association between race and neutral genetic variation should persist in samples collected well away from the zone but should have decayed closer to the zone, which is exactly what we observe at the SNPs differing in allele frequency across the zone (SNPs 3 and 5 at Tende; 4, 6 and 7 at Seyne): the cline width, as estimated from the inverse of the slope of allele frequency at the zone center, ranges between 4.5 and 5.3 km. Graphs of these clines are provided in the online Supplementary Appendix S1.

Selection for ancestral alleles and combinations in the zone can widen clines (see Gavrilets 1997; Shuker et al. 2005), so broad rDNA clines do not necessarily mean that the rDNA differences play no part in the genetic incompatibility between the races. However, the absence of an association between rDNA sequence and karyotype in transects across the hybrid zone (Fig. 5) suggests that rDNA differences are not implicated in the selection maintaining the much narrower transition in the karyotype. Our result is in pronounced contrast to that of Dallas et al. (1988) who detected strong linkage disequilibrium between karyotype and one particular rDNA restriction polymorphism in the Tende area (but interestingly not at Lac Autier further west). However, new FISH results can explain Dallas et al.'s results by physical linkage. In contrast to all other populations studied so far, fused individuals from Tende have an rDNA locus on the centromere of the fused X chromosome ($N = 28$; P. Veltsos, unpubl. data), which could be the location of the variant in question.

In summary, our results have made use of the extensive variation in rDNA composition in *P. pedestris.* Because rDNA is a highly repeated sequence, we can estimate the frequency of variants among the rDNA copies within an individual. Their frequency can change rapidly over the generations: sufficiently rapidly, in some cases, that there are large differences between individuals of the same population. Furthermore, there is extensive variation between populations. Over short distances (hundreds of meters) the genetic differences increase smoothly—suggesting homogenization by gene flow. Over longer distances (kilometers) the differences are more erratic and reflect the known history of colonization of the area. They also reveal previously unappreciated patterns of colonization. This interpretation suggests that, although the variation is generated rapidly, it can persist for thousands of generations. This conclusion is consistent with our previous analysis of rDNA sequence variants, which suggested the persistence of rDNA variants over even longer time periods (millions of generations; Keller et al. 2006).

ACKNOWLEDGMENTS

We are very grateful to M. Quinlivan for his invaluable help with establishing the pyrosequencing assay. We thank C. Mein, R. Joseph, and C. Fleischmann for support in the laboratory and I. Chintauan-Marquier, A. Gebreselassie, H. Jahan, A. Leitch, E. Leitch, and J. Nichols for assistance in the field and helpful discussions. This work was financially supported by fellowships from the Swiss National Science Foundation (grant PBBEA-104447) and the Roche Research Foundation to IK and a Natural Environment Research Council grant (NER-B-S-2003–00859) to RAN.

LITERATURE CITED

Álvarez, I., and J. F. Wendel. 2003. Ribosomal ITS sequences and plant phylogenetic inference. Mol. Phylogenet. Evol. 29:417–434.

Barton, N. H. 1979. Gene flow past a cline. Heredity 43:333–339.

———. 1983. Multilocus clines. Evolution 37:454–471.

Barton, N. H., and G. M. Hewitt. 1981a. A chromosomal cline in the grasshopper *Podisma pedestris*. Evolution 35:1008–1018.

———. 1981b. The genetic basis of hybrid inviability in the grasshopper *Podisma pedestris*. Heredity 47:367–383.

———. 1982. A measurement of dispersal in the grasshopper *Podisma pedestris*. Heredity 48:237–249.

———. 1985. Analysis of hybrid zones. Annu. Rev. Ecol. Syst. 16:113–148.

Baur, B., H. Baur, C. Roesti, and D. Roesti. 2006. Die Heuschrecken der Schweiz. Haupt Verlag, Bern.

Bella, J. L., M. Westerman, C. Lopez-Fernandez, J. Delatorre, J. M. Rubio, and J. Gosalvez. 1991. Sex chromosome and autosome divergence in *Podisma* (Orthoptera) in western Europe. Genet. Sel. Evol. 23:5–13.

Crease, T. J. 1995. Ribosomal DNA evolution at the population level: nucleotide variation in intergenic spacer arrays of *Daphnia pulex*. Genetics 141:1327–1337.

Dallas, J. F., N. H. Barton, and G. A. Dover. 1988. Interracial rDNA variation in the grasshopper, *Podisma pedestris*. Mol. Biol. Evol. 5:660–674.

Eickbush, T. H., and D. G. Eickbush. 2007. Finely orchestrated movements: evolution of the ribosomal RNA genes. Genetics 175:477–485.

Elder, J. F., and B. J. Turner. 1995. Concerted evolution of repetitive DNA sequences in eukaryotes. Q. Rev. Biol. 70:297–320.

Flanagan, N. S., P. L. Mason, J. Gosalvez, and G. M. Hewitt. 1999. Chromosomal differentiation through an Alpine hybrid zone in the grasshopper *Chorthippus parallelus*. J. Evol. Biol. 12:577–585.

Gavrilets, S. 1997. Hybrid zones with Dobzhansky-type epistatic selection. Evolution 51:1027–1035.

Gosalvez, J., J. L. Bella, and G. M. Hewitt. 1988. Chromosomal differentiation in *Podisma pedestris*: a third race. Heredity 61:149–157.

Graur, D., and W. H. Li. 2000. Fundamentals of molecular evolution. Sinauer Associates, Sunderland, MA.

Halliday, R. B., N. H. Barton, and G. M. Hewitt. 1983. Electrophoretic analysis of a chromosomal hybrid zone in the grasshopper *Podisma pedestris*. Biol. J. Linn. Soc. 19:51–62.

Halliday, R. B., S. F. Webb, and G. M. Hewitt. 1984. Genetic and chromosomal polymorphism in hybridizing populations of the grasshopper *Podisma pedestris*. Biol. J. Linn. Soc. 21:299–305.

Hewitt, G. M., and B. John. 1972. Inter-population sex chromosome polymorphisms in the grasshopper *Podisma pedestris*. Chromosoma 37: 23–42.

Keller, I., I. C. Chintauan-Marquier, P. Veltsos, and R. A. Nichols. 2006. Ribosomal DNA in the grasshopper *Podisma pedestris*: escape from concerted evolution. Genetics 174:863–874.

Kolak, M., F. Karpati, H. J. Monstein, and J. Jonasson. 2003. Molecular typing of bacterial flora in sputum of cystic fibrosis patients. Int. J. Med. Microbiol. 293:309–317.

Kovarik, A., R. Matyasek, K. Y. Lim, K. Skalicka, B. Koukalova, S. Knapp, M. Chase, and A. R. Leitch. 2004. Concerted evolution of 18–5.8–26S rDNA repeats in *Nicotiana* allotetraploids. Biol. J. Linn. Soc. 82:615–625.

Lavebratt, C., S. Sengul, M. Jansson, and M. Schalling. 2004. Pyrosequencing^{TM} based SNP allele frequency estimation in DNA pools. Hum. Mutat. 23:92–97.

Liao, D. 1999. Concerted evolution: molecular mechanism and biological implications. Am. J. Hum. Genet. 64:24–30.

Monstein, H. J., S. Nikpour-Badr, and J. Jonasson. 2001. Rapid molecular identification and subtyping of *Helicobacter pylori* by pyrosequencing of the 16S rDNA variable V1 and V3 regions. FEMS Microbiol. Lett. 199:103–107.

Nichols, R. A., and G. M. Hewitt. 1986. Population structure and the shape of a chromosomal cline between two races of *Podisma pedestris* (Orthoptera: Acrididae). Biol. J. Linn. Soc. 29:301–316.

Nichols, R. A., and G. M. Hewitt. 1988. Genetical and ecological differentiation across a hybrid zone. Ecol. Entomol. 13:39–49.

Nichols, R.A., E. A. Humpage, and G. M. Hewitt. 1990. Gene flow and the distribution of karyotypes in the alpine grasshopper *Podisma pedestris* (L.) (Orthoptera: Acrididae). Boletin de Sanidad Vegetal Plagas (Fuera de Serie) 20:373–379.

Nieto Feliner, G., B. Gutierrez Larena, and J. Fuertes Aguilar. 2004. Fine-scale geographical structure, intra-individual polymorphism and recombination in nuclear ribosomal internal transcribed spacers in *Armeria* (Plumbaginaceae). Ann. Bot. 93:189–200.

Ohta, T., and G. A. Dover. 1983. Population genetics of multigene families that are dispersed into two or more chromosomes. Proc. Natl. Acad. Sci. USA 80:4079–4083.

Parkin, E. J., and R. K. Butlin. 2004. Within- and between-individual sequence variation among ITS1 copies in the meadow grasshopper *Chorthippus parallelus* indicates frequent intrachromosomal gene conversion. Mol. Biol. Evol. 21:1595–1601.

Rauscher, J. T., J. J. Doyle, and A. H. D. Brown. 2002. Internal transcribed spacer repeat-specific primers and the analysis of hybridization in the *Glycine tomentella* (Leguminosae) polyploid complex. Mol. Ecol. 11:2691–2702.

R Development Core Team. 2007. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org.

Sharpe, R. G., R. E. Harbach, and R. Butlin. 2000. Molecular variation and phylogeny of members of the Minimus group of *Anopheles* subgenus *Cellia* (Diptera: Culicidae). Syst. Entomol. 25:263–272.

Shuker, D. M., K. Underwood, T. M. King, and R. K. Butlin. 2005. Patterns of male sterility in a grasshopper hybrid zone imply accumulation of hybrid incompatibilities without selection. Proc. R. Soc. B 272:2491–1497.

Sokal, R. R., and F. J. Rohlf. 1995. Biometry. W. H. Freeman and Company, New York.

Tarnberg, M., T. Jakobsson, J. Jonasson, and U. Forsum. 2002. Identification of randomly selected colonies of lactobacilli from normal vaginal fluid by pyrosequencing of the 16S rDNA variable V1 and V3 regions. APMIS 110:802–810.

Westerman, M., N. H. Barton, and G. M. Hewitt. 1987. Differences in DNA content between two chromosomal races of the grasshopper *Podisma pedestris*. Heredity 58:221–228.

Westerman, M., and G. M. Hewitt. 1985. Chromosome banding in *Podisma pedestris*. Heredity 55:157–161.

Associate Editor: J. Feder

Supplementary Material

The following supplementary material is available for this article:

Appendix S1. Pyrosequencing data for 175 individuals of the grasshopper *P. pedestris* genotyped at seven SNP loci.

This material is available as part of the online article from:
http://www.blackwell-synergy.com/doi/abs/10.1111/j.1558-5646.2007.00320.x
(This link will take you to the article abstract).

Please note: Blackwell Publishing is not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.