Demography and weak selection drive patterns of transposable element diversity in natural populations of *Arabidopsis lyrata*

Steven Lockton, Jeffrey Ross-Ibarra, and Brandon S. Gaut*

Department of Ecology and Evolutionary Biology, University of California, Irvine, CA 92697

Edited by M. T. Clegg, University of California, Irvine, CA, and approved June 25, 2008 (received for review May 13, 2008)

Transposable elements (TEs) are the major component of most plant genomes, and characterizing their population dynamics is key to understanding plant genome complexity. Yet there have been few studies of TE population genetics in plant systems. To study the roles of selection, transposition, and demography in shaping TE population diversity, we generated a polymorphism dataset for six TE families in four populations of the flowering plant *Arabidopsis lyrata*. The TE data indicated significant differentiation among populations, and maximum likelihood procedures suggested weak selection. For strongly bottlenecked populations, the observed TE band-frequency spectra fit data simulated under neutral demographic models constructed from nucleotide polymorphism data. Overall, we propose that TEs are subjected to weak selection, the efficacy of which varies as a function of demographic factors. Thus, demographic effects could be a major factor driving distributions of TEs among plant lineages.

bottleneck | genetics | TE-display

Transposable elements (TEs) are a major component of plant genomes, comprising >50% of all large (>2,000 Mb) angiosperm genomes studied to date (1). In the 2,500-Mb maize genome, for example, TE amplification is the source of 60%–80% of the genomic sequence (2, 3). TEs are also abundant in the compact genomes of rice (430 Mb) and *Arabidopsis thaliana* (119 Mb), contributing ≈29% and ≈10% of their genomes, respectively (2, 4). Because the mean haploid angiosperm genome size is ≈6,400 Mb (5), it is no exaggeration to state that most of the DNA contained within the nuclei of flowering plants is, in fact, TE DNA. If one is to understand the dynamics and evolution of plant genomes, a comprehensive understanding of TE evolutionary dynamics is therefore necessary.

Population genetics is a powerful tool with which to study the evolutionary dynamics of TEs. TE population genetics has a particularly rich empirical history in *Drosophila melanogaster*, in which the surprisingly few occupied TE sites are found at low population frequencies (6, 7). What limits the number and population frequencies of TEs? Most models of TE population dynamics have focused on the maintenance of TE copy number via an equilibrium between transposition, which increases the abundance of TEs in a host genome, and natural selection, which removes deleterious TE insertions (8, 9). However, the number and distribution of TEs in genomes are unlikely to be determined by selection and transposition alone (10); factors such as the population and life history of the host may also play significant roles (11). Several recent studies of nucleotide polymorphism highlight the difficulty of identifying the signature of natural selection without first understanding the impact of demographic history (12–16). Yet the role of population structure in shaping TE distributions has been largely unexplored at empirical and theoretical levels (17). Without data on TE abundance within and among natural plant populations, our understanding of the evolutionary forces shaping TE distributions will remain incomplete.

Here we study the population genetics of TEs in *A. lyrata* (18). Much is known about TEs in the genus *Arabidopsis*, because the approximately 6,000 TEs within the *A. thaliana* genome have been well characterized (4, 19). *A. lyrata* diverged from *A. thaliana* ≈5 million years ago (20) and has become a model system for plant molecular population genetics (21). *A. lyrata* is a predominantly self-incompatible, perennial species distributed across northern and central Europe, Asia, and North America. *A. lyrata* consists of large, stable populations, particularly in Central Europe where populations are hypothesized to have served as Pleistocene refugia (21–23). Importantly, Ross-Ibarra *et al.* (24) modeled the demographic history of six natural *A. lyrata* populations based on single-nucleotide polymorphism (SNP) data from 77 nuclear genes. They compared a putatively refugial German population to five populations from Sweden, Iceland, Russia, the United States, and Canada, estimating divergence times and population size differences between the German and non-German populations. The non-German populations had from ≈7- to 18-fold smaller estimated effective population sizes (N_e) than German populations, consistent with bottlenecks during colonization from Central European refugia (25).

Although much is known about the molecular population genetics of *A. lyrata*, very little is known about the population genetics of TEs in this or any other plant. Sampling five of the same populations studied by Ross-Ibarra *et al.* (24), we use transposon-insertion display (hereafter referred to as TE display) (26, 27) to generate TE polymorphism datasets for members of six TE families. With this large dataset, we exploit the inferred demographic history of *A. lyrata* to characterize the evolutionary forces that act on TEs at the population level.

Results

TE-Display Data and Population Frequencies. We performed TE display in members of six TE families: *Gypsy*-like (*Gypsy*) class I LTR-retrotransposons; SINE-like I (SINE) and LINE-like (LINE) class I non-LTR-retroelements; and *Ac*-like III (*Ac*), CACTA-like (CACTA), and *Tourist*-like MITE (MITE) class II DNA elements. TE display was applied to samples of 9 to 12 individuals from each of four natural populations of *A. lyrata*: the putatively refugial German population, the colonized Swedish and Russian populations, and a combined North American sample from Canadian and U.S. populations (Table 1).

Fig. 1 graphically represents the *Ac* diversity data; analogous

Author contributions: S.L. and B.S.G. designed research; S.L. performed research; S.L. and J.R.-I. analyzed data; and S.L., J.R.-I., and B.S.G. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. EU558519–EU558534).

*To whom correspondence should be addressed at: 321 Steinhaus Hall, University of California, Irvine, Irvine, CA 92697-2525. E-mail: bgaut@uci.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0804671105/DCSupplemental.

Table 1. Descriptive statistics for TE bands

TE family	Mean polymorphic band frequency	var(f)*	S†	Sx‡	Sf§
Germany (n = 11)					
Gypsy	0.36	0.11	17	9.0	3
LINE	0.36	0.07	20	10.7	0
SINE	0.35	0.08	30	11.7	2
Ac	0.35	0.06	25	8.3	2
CACTA	0.26	0.05	17	12.7	0
MITE	0.23	0.03	48	29.3	1
North America (n = 12)					
Gypsy	0.48	0.12	13	5	5
LINE	0.31	0.07	20	8	4
SINE	0.51	0.10	25	4	5
Ac	0.32	0.07	37	19	1
CACTA	0.35	0.10	12	7	0
MITE	0.27	0.04	22	8	4
Russia (n = 12)					
Gypsy	0.47	0.09	12	3	6
LINE	0.25	0.04	25	15	4
SINE	0.41	0.10	19	1	4
Ac	0.45	0.07	22	9	2
CACTA	0.35	0.08	10	7	0
MITE	0.47	0.11	30	11	1
Sweden (n = 9)					
Gypsy	0.49	0.11	11	4	2
LINE	0.43	0.07	14	5	1
SINE	0.43	0.06	21	4	7
Ac	0.46	0.07	30	13	2
CACTA	0.30	0.04	12	7	1
MITE	0.44	0.10	33	10	5

*Variance of polymorphic band frequencies.
†Number of observed TE bands, ignoring species-wide fixed bands.
‡Number of unique TE bands in a pairwise comparison between Germany and each other population. For Germany, a mean from each pairwise comparison is shown.
§Numbers of within-population fixed TE bands, ignoring species-wide fixed bands.

figures for the other five TE families are available [supporting information (SI) Fig. S1]. The TE band data yield three initial observations. First, an appreciable proportion of TE bands are fixed within population samples, but fixed bands make up a smaller proportion of total diversity in the German population (Table 1). Second, each polymorphic TE band is found, on average, in multiple individuals. For example, estimates of the mean within-population band frequency ($\bar{f}_w$) for *Ac* range from 0.32 in the North American sample to 0.46 in the Swedish population, with similar ranges of $\bar{f}_w$ for the other five element families (Table 1). Note also that $\bar{f}_w$ and the variance of f_w are often lowest in the German population (Table 1). Finally, although most bands are found in multiple populations, each TE family yields unique bands in every population. For example, 16 of 54 (30%) observed *Ac* bands are unique to one of the four population samples, with just two of these specific to the German sample (Fig. 1).

Population Differentiation. To investigate the extent of population differentiation, we applied a molecular analysis of variance (AMOVA) (28) to data from each TE family. Permutation tests revealed significant population differentiation (as measured by the Φ_{PT} statistic) for each pairwise population comparison for all TE families (Fig. S2) ($\Phi_{PT} > 0$ at $P < 0.01$). Comparisons that included the German sample had lower Φ_{PT} values on average (Fig. S2). Additionally, using the *Structure* program, we performed analyses using all of the TE-display data as a single dataset and assuming the TE bands were unlinked (29). A model of $K = 4$ clusters yielded the highest likelihood, with clear separation of individuals by geographic origin (Fig. S2).

Maximum Likelihood Estimation of Selection. To infer the strength of selective forces acting on TE insertions, we applied a modification of the diffusion-approximation approach of Petrov *et al.* (6), correcting for our ascertainment scheme and assuming Hardy–Weinberg equilibrium (HWE) (see *SI Methods*). We obtained the maximum likelihood estimate (MLE) of s for each TE family in each population and calculated the estimated $N_e s$ ($N_e \hat{s}$) for each (Table S1). Eight of twenty-four $N_e \hat{s}$ values have an absolute value <1, and two-thirds of the point estimates are positive. However, only 3 of 24 have confidence intervals that do not overlap zero, and only one of these is negative (MITEs in Germany). Importantly, the sign and magnitude of $N_e \hat{s}$ vary by population: When data from all elements are combined, only the German sample yields a negative $N_e s$ estimate [$N_e \hat{s} = -0.612$; 95% C.I. (-1.360, 0.289)], whereas the other three bottlenecked populations yield positive values [North America = 0.558 (-0.491, 2.850); Russia = 0.720 (-0.417, 3.612); and Sweden = 1.662 (0.072, >6.0)]. These observations raise the possibility that $N_e \hat{s}$ values reflect properties of populations as much as properties of selection on TEs.

Population Bottleneck TE Dynamics. To assess the relative contributions of transposition, selection, and demography on patterns of TE diversity, for each TE family we compared the German population (as a reference) to each of the other populations in turn. We focused on three summary statistics of the data (Fig. 2): the total number of bands in the two populations (S_{tot}), the number of unique bands in the bottlenecked population (Sx_b), and the total number of bands in the bottlenecked population (S_b). Of particular note is the fact that Sx_b was higher in some non-German populations than one might intuitively expect in bottlenecked populations; for example, $Sx_b = 19$ for *Ac* elements in the bottlenecked North American population compared with, on average, approximately eight *Ac* bands unique to Germany in pairwise comparisons (Table 1).

We compared S_b and Sx_b to demographic expectations by using simulations of models that were inferred from silent nucleotide polymorphisms (24) (Fig. 2) (see *Methods*). These simulations were conditioned on S_{tot} and assumed a constant (but unknown) transposition rate. We found significantly higher pairwise Sx_b than expected in nearly a third (5 of 18) of our comparisons (North America: *Ac*, CACTA, LINE; Russia: LINE; Sweden: *Ac*; $P < 0.05$) (data not shown). Three scenarios could explain this observation. First, the demographic model used may be incorrect. This possibility seems unlikely, as the same model fits SNP data from the same populations well (24). Second, transposition could lead to higher Sx_b than demographic expectations. However, the number of total and unique TE bands in Germany fit simulations well for every pairwise comparison (data not shown). Thus, this explanation only makes sense if transposition rates are substantially higher in the bottlenecked populations relative to Germany. Finally, an excess of Sx_b could be explained by purifying selection removing TEs in the German population, thus increasing the number of TEs appearing as unique to other populations.

To investigate this last possibility, we performed simulations across a grid of θ values for the German population ($\theta = 4N_e\mu$, where μ is the population transposition rate). We simulated data from models with θ values decreasing from 100% to 10% of the original value for Germany (24) (Fig. 3). In this context, decreasing θ serves as a proxy for weak selection (30). For 14 of 18 combinations (i.e., six TE families × three comparison populations), the observed data better fit a model with decreased θ (Fig. 3). Data from *Ac*, LINE, and CACTA elements were particularly compelling, with at least 10-fold estimated decreases of θ in Germany. These results suggest that purifying selection acts on TEs in the German population relative to a null demographic model fitted with presumably neutral SNP polymorphisms.

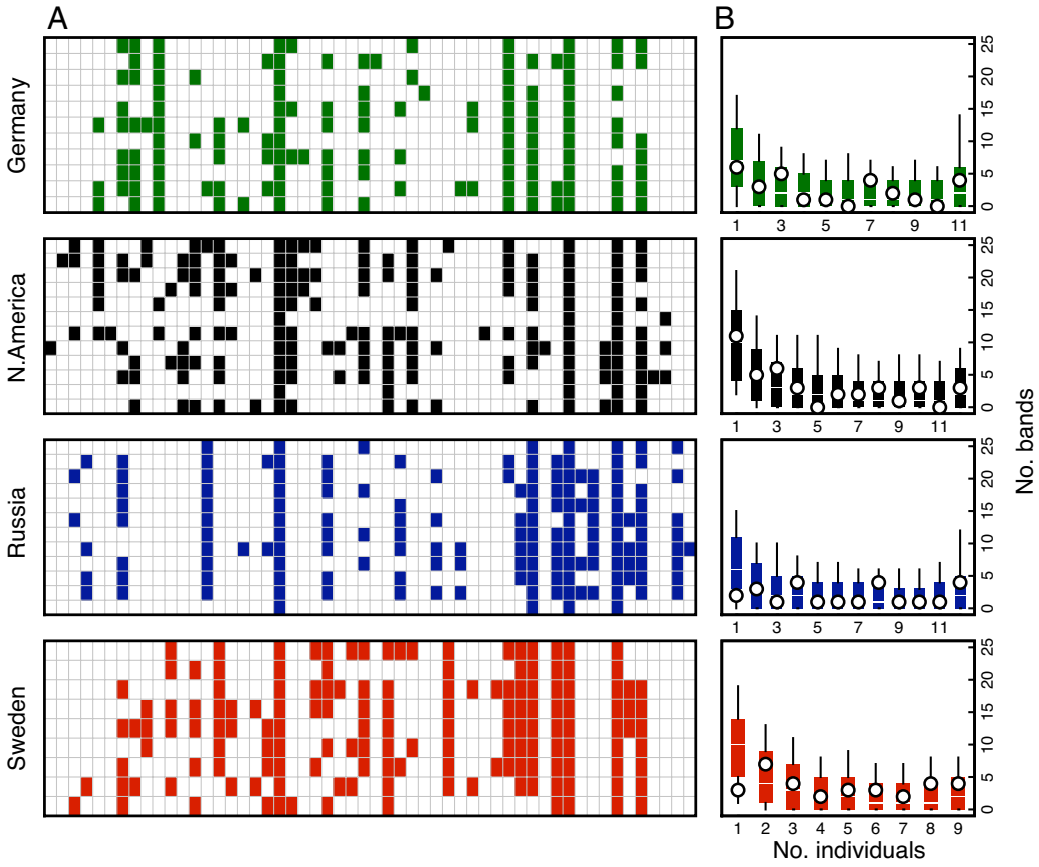

Fig. 1. *A. lyrata Ac*-like diversity data. (*A*) A plot of TE-display data. A colored cell represents the presence of a TE; a white cell is a lack of TE detection. Each column represents a TE band, and each row represents an individual. Colors show population of origin. (*B*) TE BFS for observed data (circles) and simulated data (bars and vertical black lines). The bars represent the 95% credible intervals, the white horizontal lines in the bars are the medians, and the vertical black lines show the full ranges of the simulations.

Previous studies have noted that bottleneck events decrease N_e, thereby decreasing the efficacy of selection and slowing the rate of TE loss (31). If purifying selection is weakened due to bottleneck-related reductions in N_e then we predict that the frequency spectra of TE bands in bottlenecked populations should be consistent with neutrality under the inferred demographic history. To test this hypothesis, we again used the demographic models of Ross-Ibarra *et al.* (24), simulating data under the most likely value of θ (Fig. 3)

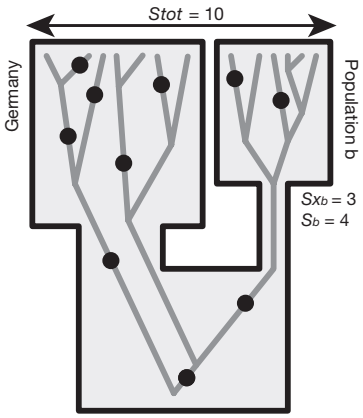

Fig. 2. Hypothetical genealogy to illustrate the simulation of TE distributions under a neutral bottleneck model. Black circles represent transposition events.

to generate a posterior probability distribution of band-frequency spectra (BFS) for neutral multilocus TE-display datasets (see *Methods*). The observed BFS from our TE-display data fit the neutral expectations from these simulations quite well (Fig. 1, Fig. S1, and Table S2). Across all families and all populations, just 3.0% (8 of 265) of observed frequencies fell outside the 95% confidence intervals, and none remained significant after a per-TE family Bonferroni correction (Fig. 1, Fig. S1, and Table S2).

In an effort to bolster statistical power, we pooled data across TE families and compared the observed (pooled) BFS to its expectation based on simulation. To make these comparisons, for each population we first calculated the Mann–Whitney U between the observed pooled BFS and 1,000 simulated BFS, retaining the mean of these 1,000 U statistics. We then repeated this procedure for 1,000 simulated BFS, comparing each with 1,000 simulated data sets and retaining the mean U statistic. These 1,000 mean values represent variation in the pooled BFS expected under the demographic model and form the basis for evaluating the fit of the observed BFS to demographic expectations (Fig. 4). The pooled BFS did not differ from demographic expectations for the three non-German populations (Fig. 4) (North America: $P = 0.346$; Russia: $P = 0.530$; Sweden: $P = 0.708$). However, the pooled BFS in Germany was different from simulated data (Fig. 4) ($P = 0.050$), with an excess of observed low frequency variants. This last observation is consistent with weak purifying selection skewing the distribution of TEs toward rare variants in the German population.

Direct Comparison of TEs and Silent SNPs. If purifying selection against TEs is relaxed in the bottlenecked populations, comparisons

EVOLUTION

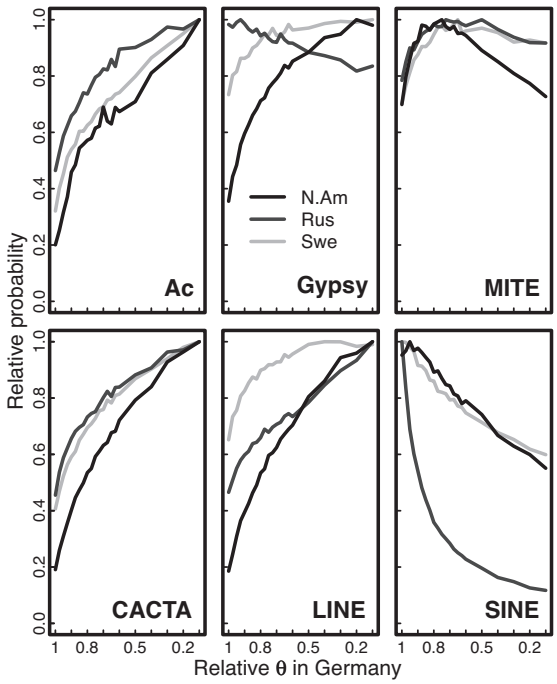

Fig. 3. Relative probabilities of the observed data for different values of θ in Germany, scaled to the original demographic model (24).

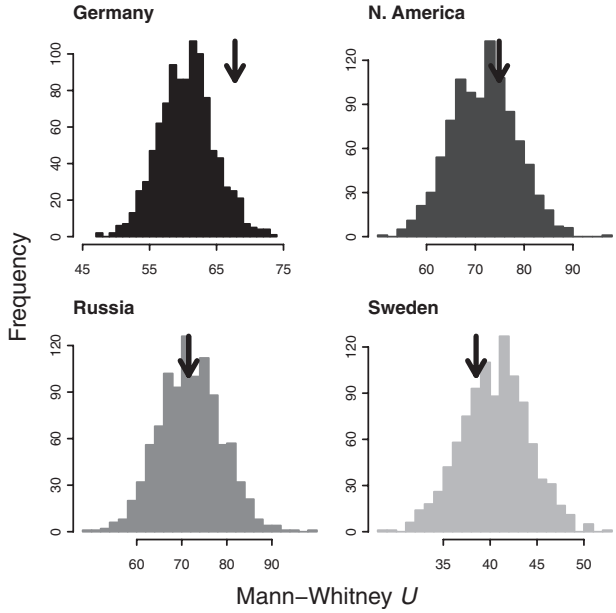

Fig. 4. Mann–Whitney *U* statistics of pooled TE-band frequency spectra. Histograms depict the distribution of 1,000 mean *U* statistics for data simulated under the neutral demographic model for each population, and arrows point to the mean *U* value obtained by comparing observed data with the model. The observed value in Germany is $P = 0.05$.

of the ratio of TE diversity to diversity at neutral SNPs should reveal differences among the populations, with higher ratios in populations with relaxed purifying selection. Indeed, this is exactly the pattern observed: In Germany, the ratio of counts of polymorphic TEs to polymorphic derived silent SNPs is 0.28, but the three other populations have ratios of 0.86, 0.75, and 0.94 (Table S3). The difference is significant for all pairwise comparisons with Germany (Fisher's Exact Test, $P < 0.001$). A similar comparison of the mean number of TE bands per individual to the mean number of polymorphic SNPs per diploid genome produces identical results, in that Germany has a lower ratio (0.15) than North America (0.45), Russia (0.45), and Sweden (0.45) (Fisher's Exact Test, $P < 0.001$ for all pairwise comparisons to Germany) (Table S3). These observations suggest that patterns of TE polymorphisms differ across populations relative to a neutral marker (silent SNPs).

Discussion

TEs represent the majority of plant genomic DNAs and undoubtedly contribute to genomic flux. Molecular evolutionary analyses suggest, for example, that many plant TEs have proliferated within the recent past (32–35) and that proliferation is counteracted by TE deletion (36, 37). Nonetheless, our understanding of TE evolution in plant genomes is woefully incomplete, in part because there have been few population genetic studies of plant TEs. Without population genetic information, one cannot infer the relative roles of transposition, natural selection, and genetic drift in TE accumulation.

Most population genetic analyses of TEs have assumed that TE population frequencies are governed by an equilibrium between selection and transposition (38). By using this assumption, negative selection has been found against TEs in both *Drosophila* (6, 39) and humans (40). However, recent simulation work strongly suggests that equilibrium is very unlikely under realistic conditions and that factors such as population-size variation can strongly affect TE dynamics (10). In plants, although several studies have used TE-display bands as genetic markers (37, 41, 42) or for phylogeographic analysis (43), the population genetic ramifications of TEs have largely been ignored. One notable exception inferred negative selection against TEs in *A. lyrata* under equilibrium conditions (44). Another recent study used *A. thaliana* TE polymorphism data to conclude that longer *Helitrons* are less likely to persist in the genome (32). Given the rarity of plant TE population genetic data, our population dataset of 6 TE families based on 44 individuals from 4 natural populations is to our knowledge unprecedented.

Individual TE bands are found in intermediate-to-high population frequencies in *A. lyrata*. This observation superficially suggests that the TEs in our sample have not been subjected to strong purifying selection. Assuming TE insertions are at HWE, the mean TE allele frequency was 0.24 across TE families and populations, and frequencies ranged from 0.13 to 0.35. Our allele frequency estimates match well with previous work on the *Ac* family of elements in *A. lyrata* (44), but are somewhat higher than those estimated in *Drosophila*. Mean allele frequencies of non-LTR elements in *Drosophila* are as high as 16% (6) (compared with 22% for *SINEs* and *LINEs* in our data), and even lower for LTR elements (compared with 26% for *Gypsy* in our data) (39). Mean frequencies of polymorphic TEs are higher in other systems, however, including *Ta1* in humans [36% (40)], class I TEs in pufferfish [43% (45)], and nonautonomous *Helitrons* in *A. thaliana* [60% (32)].

Evolutionary Forces Governing TE Polymorphism. Given the demographic history of *A. lyrata*, what forces govern TE diversity and polymorphism? Transposition may have occurred during the history of our sample, based on two lines of evidence. The first is the simple observation that every population sample has unique bands relative to the four other population samples. Given pairwise divergence time estimates (24) and assuming unique bands represent transposition events, we can use the average number of unique bands per individual (pooled across TE families) to calculate a per locus estimate of the transposition rate for each population. These estimates yield a mean rate of 2×10^{-5} bands per generation per locus, which is similar to estimates in ref. 46 or less than those in refs. 47 and 48. Second, transposition is biologically plausible, because the elements studied show evidence of activity in *A. thaliana*. For

example, CACTA TEs are active in methylation-deficient mutants of *A. thaliana* (48), some families of *Ac*-like elements show evidence of recent activity (47, 49), and both SINE-like and *Gypsy*-like elements have been inferred to be active within *A. thaliana*'s recent past (35, 50). In addition, SINE, *Ac*-like, CACTA, and MITE TEs are presumed to have been active recently because they contain ORFs (19) or vary in location among *A. thaliana* ecotypes (51).

Interpreting the selective forces acting on TEs is more difficult. Estimates of $N_e s$ were not large; 88% (21 of 24) had confidence intervals encompassing zero, suggesting the TE bands in our sample are subjected to at most weak selection. Nonetheless, point estimates of $N_e s$ tended to be positive, with two values significantly >0 (Table S1). Taken at face value, positive $N_e \hat{s}$ values suggest positive selection on TEs. We believe such a conclusion would be in error, however, because the diffusion models make demographic assumptions, such as large, constant population sizes, which may not apply to our *A. lyrata* populations. Supporting this view is the fact that bottlenecked populations (North America, Sweden, and Russia) yield overall $N_e \hat{s}$ values >1.0, whereas Germany, which most closely represents a neutral equilibrium population (24), yields the only overall negative $N_e s$ estimate (−0.612). We conclude, then, that the $N_e \hat{s}$ values are generally consistent with weak selection (i.e., $|N_e s| < 1$), but caution that our estimates of selection, and perhaps those of previous studies (6, 40, 45), should be viewed with healthy skepticism because they do not incorporate demographic complexities.

To further investigate the possibility of selection against TEs while recognizing demographic history, we compared the observed TE BFS to simulated BFS based on demographic models fitted to DNA sequence diversity data (24). We reasoned that purifying selection against TE insertions should lead to an overabundance of observed low-frequency TE bands relative to the expectation based on the neutral demographic models. By pooling data, we were able to show that the German population has an excess of low-frequency TE bands, as might be expected under weak purifying selection, an observation consistent with low but negative estimates of $N_e s$ for this population and estimated values of θ (Fig. 3). In contrast, we were unable to clearly reject the hypothesis that the TE distributions result largely from demographic instead of selective processes in the three strongly bottlenecked populations.

Implications for Understanding the Forces Acting on TE Diversity. Our data provide evidence for the geographic structure of *A. lyrata* populations and TE activity within and among populations. However, unlike other studies (6, 44, 51–54), we did not uncover clear evidence for selection against TEs in the non-German populations. Given the lack of obvious evidence for selection, can we discount selection entirely? The answer is no for three reasons. First, although we are unaware of any other empirical study that has explicitly modeled demographic history in TE population genetics, our statistical power to infer weak selection against a demographic background may be low. Second, previous studies have demonstrated convincingly that there is selection against TEs. For example, a recent study of rice TEs found that insertions into gene regions are lost rapidly because of strong selection against the interruption of gene function (55). These highly deleterious events are not expected to rise to appreciable population frequencies and are thus unlikely to have been included in our sample.

Third, there is the intriguing possibility that demography interacts with selection to shape the frequency and distribution of TEs. The pooled TEs in the German population have a negative $N_e \hat{s}$, but $N_e \hat{s}$ values are slightly positive in the bottlenecked populations. To the extent that these $N_e \hat{s}$ values are reasonable, they suggest that TE insertions are, on average, subject to nearly neutral population dynamics (56). The efficacy of selection is a function of N_e; if N_e changes such that $|N_e s| \ll 1.0$, drift can overcome selection. Our $N_e s$ estimates (Table S1), along with our observation that the BFS of TEs in bottlenecked populations are consistent with neutral demographic processes, are consistent with reduced efficacy of selection in bottlenecked populations. Ratios of polymorphic TEs and silent SNPs further suggest that purifying selection on TEs is relaxed in the bottlenecked populations relative to the German population.

If this conjecture is true, it has a profound impact on our understanding of the evolution of plant genomes. It suggests that genomic flux in TEs occurs at a rate that is influenced by demographic history. All other things being equal, plant species with small populations sizes should purge TE insertions less efficiently and hence accrue DNA more rapidly. The idea that genomic complexity is related to population size is not new (57) and has been cited as the cause of the accumulation of repetitive element insertions in the human genome (31). Thus far, however, there has been no compelling evidence for this effect within and between plant populations or between plant evolutionary lineages. Yet, given the wide range of differences in N_e among plants because of breeding system and life history, and also given evidence for strong demographic effects during processes like domestication (58, 59), our results raise the possibility that the differential expansion of TEs among plant lineages could be fundamentally a function of demographic history.

Materials and Methods

Sampling and Plant Growth. Five populations of *A. lyrata* were sampled for this study: Plech, Germany (sampled by M. Clauss, Max Planck Institute of Chemical Ecology, Jena, Germany); Karhumäki, Russia (courtesy of O. Savolainen, University of Oulu, Oulu, Finland); Stubbsand, Sweden (also courtesy of O. Savolainen); Indiana Dunes, U.S. and Ontario, Canada (both provided by B. Mable, University of Glasgow, Galsgow, UK). Plants were grown at 22°C with a 16-h day for 8 weeks. DNA was extracted by using Qiagen's DNeasy Plant Mini kit.

TE Display. We followed Le and Bureau (27) in our choice of TE-display adapters and adapter primers. TE-specific primers for the *Ac* and CACTA families were from previous studies (27, 44). Additional nested TE-specific primers were chosen by (*i*) designing a large number of primers for known TE sequences, (*ii*) performing virtual TE display in the *A. thaliana* genome, and (*iii*) screening primers in both *A. thaliana* ecotype Columbia and *A. lyrata*. Digestion, ligation, and PCR followed the methods of Le and Bureau (27) with slight modifications (*SI Methods*). Bands were sized with fragment analysis by using the software GeneMapper 4.0 on an ABI 3100 (Applied Biosystems) using a ROX-labeled MapMarker 1000 sizing standard (BioVentures) to score bands between 60- and 1,000-bp long. Preselective and selective PCR was repeated three times for each individual. Data were scored manually; a peak was scored as a TE band if it was the same base-pair size in two or more replicates. We examined the specificity and repeatability of TE display by first assessing error rates in three biological replicates of *A. thaliana* Col-0. We estimated a mean error rate of the PCR and fragment analysis at <4% across all TE families and all TE-display bands. A sample of 16 bands was cloned by using a pGEM T Easy vector (Promega), sequenced on an ABI 3130 to confirm their identity by using BLASTn (58), and submitted to GenBank. Fifteen (94%) of the bands had homology (at an e-value <1E-5) to the correct TE family; the remaining sequence matched an unannotated *A. thaliana* centromeric region.

Population Structure. AMOVA (28) was performed on TE-display band data with GenAlEx 6 (60). For the program *Structure* (29), the data were treated as a single population of unlinked loci. We performed *Structure* on the band data with 10,000 burn-in runs followed by 100,000 steps, without using population source information and assuming the possibility of admixture. Results were visualized with the program *DISTRUCT* (61).

Simulation of the Neutral Demographic Model. We used the demographic models inferred by Ross-Ibarra *et al.* (24), combining models for U.S. and Canada for the North American simulations. We simulated TE population genetic data with the program *ms* (62), drawing parameter values from the posterior distributions of the inferred models. We conditioned simulations on the total number of occupied bands S_{tot} observed in the two populations being compared; such conditioning requires only that the (unknown) rate of transposition remain constant across the genealogy. We assumed TE sites are unlinked and for each site simulated 2*n* alleles, where *n* is the number of individuals in the sample, combining alleles into *n* dominant genotypes for comparison with TE-display data.

We performed two sets of simulations; for both, data were simulated for all six TE families in each of three pairwise population comparisons (contrasting Germany to Russia, Sweden, or North America, respectively). In the first set, we

conditioned on S_{tot} and performed 100,000 multilocus simulations, comparing Sx_b in each population with the simulated value. For the second set of simulations, we varied θ in the German and ancestral populations across a grid of θ values decreasing from 100% to 10% compared with values specified in the original model. We accepted simulations that matched the observed Sx_b, recording acceptance rates and continuing until reaching 5,000 acceptances. The relative probability of each point on the grid was estimated from the acceptance rates, and the most probable value was chosen for further use. For each of the 5,000 simulations from the most probable model, we then calculated the unfolded BFS, including fixed bands, thus generating a posterior distribution and 95% credible intervals of the BFS for each TE/population combination. The German BFS were generated by pooling the Germany simulated data from each of the three pairwise comparisons.

SNP-TE Comparisons. For comparisons to numbers of polymorphic TEs in each population, we counted SNPs in 77 sequenced *A. lyrata* loci (24) and determined their ancestral state with an *A. thaliana* outgroup.

SI. A schematic representation of the bottleneck model used for parameter estimation is available in Fig. S3, and a complete list of oligonucleotide sequences used in this work can be found in Table S4.

ACKNOWLEDGMENTS. We thank N. Komarova, S. Wright, J. Hollister, and K. Thornton for assistance and discussion; R. Gaut for technical assistance; L. DeRose-Wilson for discussion; E. Thorhallsdottir, M. Clauss, O. Savolainen, and B. Mable for seed material; and three anonymous reviewers for constructive comments. This work was supported by National Science Foundation Grant DEB-0426166 (to B.S.G.).

1. Bennetzen JL (2005) Transposable elements, gene creation and genome rearrangement in flowering plants. *Curr Opin Genet Dev* 15:621–627.
2. Messing J, *et al.* (2004) Sequence composition and genome organization of maize. *Proc Natl Acad Sci USA* 101:14349–14354.
3. SanMiguel P, *et al.* (1996) Nested retrotransposons in the intergenic regions of the maize genome. *Science* 274:765–768.
4. Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408:796–815.
5. Zonneveld BJ, Leitch IJ, Bennett MD (2005) First nuclear DNA amounts in more than 300 angiosperms. *Annals of botany* 96:229–244.
6. Petrov DA, Aminetzach YT, Davis JC, Bensasson D, Hirsh AE (2003) Size matters: Non-LTR retrotransposable elements and ectopic recombination in *Drosophila*. *Mol Biol Evol* 20:880–892.
7. Montgomery E, Charlesworth B, Langley CH (1987) A test for the role of natural selection in the stabilization of transposable element copy number in a population of *Drosophila melanogaster*. *Genet Res* 49:31–41.
8. Le Rouzic A, Deceliere G (2005) Models of the population genetics of transposable elements. *Genet Res* 85:171–181.
9. Charlesworth B, Charlesworth D (1983) The population dynamics of transposable elements. *Genet Res* 42:1–27.
10. Le Rouzic A, Boutin TS, Capy P (2007) Long-term evolution of transposable elements. *Proc Natl Acad Sci USA* 104:19375–19380.
11. Picot S, *et al.* (2008) The mariner transposable element in natural populations of *Drosophila simulans*. *Heredity* 101:53–59.
12. Tenaillon MI, U'Ren J, Tenaillon O, Gaut BS (2004) Selection versus demography: A multilocus investigation of the domestication process in maize. *Mol Biol Evol* 21:1214–1225.
13. Haddrill PR, Thornton KR, Charlesworth B, Andolfatto P (2005) Multilocus patterns of nucleotide variability and the demographic and selection history of *Drosophila melanogaster* populations. *Genome Res* 15:790–799.
14. Thornton K, Andolfatto P (2006) Approximate Bayesian inference reveals evidence for a recent, severe bottleneck in a Netherlands population of *Drosophila melanogaster*. *Genetics* 172:1607–1619.
15. Voight BF, *et al.* (2005) Interrogating multiple aspects of variation in a full resequencing data set to infer human population size changes. *Proc Natl Acad Sci USA* 102:18508–18513.
16. Wright SI, *et al.* (2005) The effects of artificial selection on the maize genome. *Science* 308:1310–1314.
17. Deceliere G, Charles S, Biemont C (2005) The dynamics of transposable elements in structured populations. *Genetics* 169:467–474.
18. Savolainen O, Langley CH, Lazzaro BP, Fr H (2000) Contrasting patterns of nucleotide polymorphism at the alcohol dehydrogenase locus in the outcrossing *Arabidopsis lyrata* and the selfing *Arabidopsis thaliana*. *Mol Biol Evol* 17:645–655.
19. Le QH, Wright S, Yu ZH, Bureau T (2000) Transposon diversity in *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* 97:7376–7381.
20. Koch MA, Haubold B, Mitchell-Olds T (2000) Comparative evolutionary analysis of the chalcone synthase and alcohol dehydrogenase loci among different lineages of *Arabidopsis*, *Arabis* and related genera (Brassicaceae). *Mol Biol Evol* 17:1483–1498.
21. Mitchell-Olds T (2001) *Arabidopsis thaliana* and its wild relatives: A model system for ecology and evolution. *Trends Ecol Evol* 16:693–700.
22. Clauss MJ, Mitchell-Olds T (2006) Population genetic structure of *Arabidopsis lyrata in Europe*. *Mol Ecol* 15:2753–2766.
23. Koch MA, Matschinger M (2007) Evolution and genetic differentiation among relatives of *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* 104:6272–6277.
24. Ross-Ibarra J, *et al.* (2008) Patterns of polymorphism and demographic history in natural populations of *Arabidopsis lyrata*. *PLoS ONE* 3:e2411.
25. Muller MH, Leppala J, Savolainen O (2008) Genome-wide effects of postglacial colonization in *Arabidopsis lyrata*. *Heredity* 100:47–58.
26. Korswagen HC, Durbin RM, Smits MT, Plasterk RH (1996) Transposon Tc1-derived, sequence-tagged sites in *Caenorhabditis elegans* as markers for gene mapping. *Proc Natl Acad Sci USA* 93:14680–14685.
27. Le QH, Bureau T (2004) Prediction and quality assessment of transposon insertion display data. *Biotechniques* 36:222–228.
28. Excoffier L, Smouse P, Quattro J (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics* 131:479–491.
29. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959.
30. Charlesworth D, Charlesworth B, Morgan MT (1995) The pattern of neutral molecular variation under the background selection model. *Genetics* 141:1619–1632.
31. Gherman A, *et al.* (2007) Population bottlenecks as a potential major shaping force of human genome architecture. *PLoS Genet* 3:e119.
32. Hollister JD, Gaut BS (2007) Population and evolutionary dynamics of helitron transposable elements in *Arabidopsis thaliana*. *Mol Biol Evol* 24:2515–2524.
33. Ma J, Devos KM, Bennetzen JL (2004) Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice. *Genome Res* 14:860–869.
34. SanMiguel P, Gaut BS, Tikhonov A, Nakajima Y, Bennetzen JL (1998) The paleontology of intergene retrotransposons of maize. *Nat Genet* 20:43–45.
35. Pereira V (2004) Insertion bias and purifying selection of retrotransposons in the *Arabidopsis thaliana* genome. *Genome Biol* 5:R79.
36. Devos KM, Brown JKM, Bennetzen JL (2002) Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. *Genome Research* 12:1075–1079.
37. Vitte C, Panaud O, Quesneville H (2007) LTR retrotransposons in rice (Oryza sativa, L.): Recent burst amplifications followed by rapid DNA loss. *BMC Genomics* 8:218.
38. Bergman CM, Bensasson D (2007) Recent LTR retrotransposon insertion contrasts with waves of non-LTR insertion since speciation in *Drosophila melanogaster*. *Proc Natl Acad Sci USA* 104:11340–11345.
39. Charlesworth B, Langley CH (1989) The population genetics of *Drosophila* transposable elements. *Ann Rev Genet* 23:251–287.
40. Boissinot S, Davis J, Entezam A, Petrov D, Furano AV (2006) Fitness cost of LINE-1 (L1) activity in humans. *Proc Natl Acad Sci USA* 103:9590–9594.
41. De Keukeleire P, *et al.* (2001) Analysis by transposon display of the behavior of the dTph1 element family during ontogeny and inbreeding of *Petunia hybrida*. *Mol Genet Genomics* 265:72–81.
42. Kwon SJ, Park KC, Kim JH, Lee JK, Kim NS (2005) Rim 2/Hipa CACTA transposon display: A new genetic marker technique in *Oryza species*. *BMC genetics* 6:15.
43. Cornman RS, Arnold ML (2007) Phylogeography of *Iris missouriensis* (Iridaceae) based on nuclear and chloroplast markers. *Mol Ecol* 16:4585–4598.
44. Wright SI, Quang HL, Schoen DJ, Bureau TE (2001) Population dynamics of an *Ac*-like transposable element in self- and cross-pollinating *Arabidopsis*. *Genetics* 158:1279–1288.
45. Neafsey DE, Blumenstiel JP, Hartl DL (2004) Different regulatory mechanisms underlie similar transposable element profiles in pufferfish and fruitflies. *Mol Biol Evol* 21:2310–2318.
46. Nuzhdin SV (1999) Sure facts, speculations, and open questions about the evolution of transposable element copy number. *Genetica* 107:129–137.
47. Frank MJ, Liu D, Tsay YF, Ustach C, Crawford NM (1997) Tag1 is an autonomous transposable element that shows somatic excision in both *Arabidopsis* and tobacco. *Plant Cell* 9:1745–1756.
48. Miura A, *et al.* (2001) Mobilization of transposons by a mutation abolishing full DNA methylation in *Arabidopsis*. *Nature* 411:212–214.
49. Tsay YF, Frank MJ, Page T, Dean C, Crawford NM (1993) Identification of a mobile endogenous transposon in *Arabidopsis thaliana*. *Science* 260:342–344.
50. Lenoir A, *et al.* (2001) The evolutionary origin and genomic organization of SINEs in *Arabidopsis thaliana*. *Mol Biol Evol* 18:2315–2322.
51. Casacuberta E, Casacuberta JM, Puigdomenech P, Monfort A (1998) Presence of miniature inverted-repeat transposable elements (MITEs) in the genome of *Arabidopsis thaliana*: Characterisation of the emigrant family of elements. *Plant J* 16:79–85.
52. Rizzon C, Marais G, Gouy M, Biemont C (2002) Recombination rate and the distribution of transposable elements in the *Drosophila melanogaster* genome. *Genome Research* 12:400–407.
53. Hoogland C, Biemont C (1996) Chromosomal distribution of transposable elements in *Drosophila melanogaster*: Test of the ectopic recombination model for maintenance of insertion site number. *Genetics* 144:197–204.
54. Wright SI, Agrawal N, Bureau TE (2003) Effects of recombination rate and gene density on transposable element distributions in *Arabidopsis thaliana*. *Genome Res* 13:1897–1903.
55. Naito K, *et al.* (2006) Dramatic amplification of a rice transposable element during recent domestication. *Proc Natl Acad Sci USA* 103:17620–17625.
56. Ohta T (1992) The nearly neutral theory of molecular evolution. *Ann Rev Syst Ecol* 23:263–286.
57. Lynch M, Conery JS (2003) The origins of genome complexity. *Science* 302:1401–1404.
58. Doebley J (1989) in *Isozymes in Plant Biology*, eds Soltis D, Soltis P (Dioscorides Press, Portland, Oregon), pp 165–191.
59. Eyre-Walker A, Gaut RL, Hilton H, Feldman DL, Gaut BS (1998) Investigation of the bottleneck leading to the domestication of maize. *Proc Natl Acad Sci USA* 95:4441–4446.
60. Peakall R, Smouse PE (2006) GENALEX 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Mol Ecol Notes* 6:288–295.
61. Rosenberg NA (2004) DISTRUCT: A program for the graphical display of population structure. *Mol Ecol Notes* 4:137–138.
62. Hudson RR (2002) Generating samples under a Wright–Fisher neutral model of genetic variation. *Bioinformatics* 18:337–338.

Supporting Information

Lockton *et al.* 10.1073/pnas.0804671105

SI Methods

Digestion, Ligation, and PCR for TE Display. We digested 5 ng of genomic DNA of each individual with 5 units of *Bfa*I following Le and Bureau (1). We diluted the subsequent reaction mix to 200 μl for use as the template in subsequent PCRs. All oligonucleotide sequences can be found in Table S4. The first, preselective round of PCR consisted of 1× buffer, 1 μM primer AP1, 1 μM TE primer 1, 0.8 mM dNTPs, 0.4 units of GeneChoice Taq, and 1.5 μl of ligation mix in a 10-μl reaction. Touchdown PCR was performed by using an ABI GeneAmp 9700 thermocycler (Applied Biosystems) at 95°C for 3 min, then 15 cycles of 95°C for 30 s, 58°C for 45 s (with each cycle, the annealing temperature was decreased by 0.5°C), and 70°C for 90 s. The PCR was completed with 10 further cycles with a static 51°C annealing temperature and a final 7-min extension at 70°C. This PCR was then diluted 100-fold in water. The second, selective PCR was identical, apart from the use of 1.5 μl of diluted preselective PCR mix as a template, 0.2 μM FAM-labeled adapter-specific primer, and the second, nested TE-specific primer. Three technical replicates were performed for each *A. lyrata* individual. Technical replicates used common genomic DNA ligations but different amplifications and fragment analysis.

Modification of the Diffusion Model to Estimate $N_e s$. We use a modification of the diffusion approach of Petrov *et al.* (2) to estimate the selection coefficient s for an element at frequency x, where $0 \le x \le 1$. Given an estimate of the effective number of diploid individuals N, the approach further assumes an infinite number of insertion sites and that the fitnesses of homozygote nulls, heterozygotes, and homozygote insertions are of 1, $1 + hs$, and $1 + s$, respectively, for $s > -1$ and dominance term $0 \le h \le 1$. Modified to include an inbreeding statistic f (3), the drift and diffusion terms for a Wright–Fisher model are

$$m[x, s, N, h] = 2Ns(x(1-x)(h + x(1-2h) + f(1 - x - h + 2xh))) \quad [1]$$

and

$$v[x] = x(1-x). \quad [2]$$

Following Petrov *et al.* (2), and assuming elements are introduced to the population at frequency $1/2N$, we let

$$\bar{\tau}[x, s, N, h] = \frac{2}{v[x]\psi[x, s, N, h]g[0, 1]} \left(g\left[\frac{1}{2N}, 1\right]g[0, x]\theta\left[\frac{1}{2N} - x\right] + g\left[0, \frac{1}{2N}\right]g[x, 1]\theta\left[x - \frac{1}{2N}\right]\right), \quad [3]$$

where

$$\theta[z] = \begin{cases} 1, & z > 0 \\ \frac{1}{2}, & z = 0 \\ 0, & z < 0 \end{cases} \quad [4]$$

$$\psi[x, s, N, h] = e^{-2\int_0^1 \frac{m[x,s,N,h]}{v[x]}dx}, \quad [5]$$

and

$$g[a, b] = \int_a^b \psi[x, s, N, h]dx. \quad [6]$$

Assuming a large number of elements in a population at transposition-selection equilibrium, the frequency spectrum of elements is approximated by

$$F[x, s, N, h] = \frac{\bar{\tau}[x, s, N, h]}{\int_0^1 \bar{\tau}[x, s, N, h]dx}. \quad [7]$$

The probability of observing i of k diploid individuals with an insertion at a given site, using dominant data and correcting for sampling under Hardy–Weinberg equilibrium, is then

$$P[i, s, N, h] = \binom{k}{i}\int_0^1 F[x](1 - (1-x)^2)^i((1-x)^2)^{k-i}dx. \quad [8]$$

Because we can observe only a small subset of all of the elements for which $i = 0$, we ignore this class of elements. Similarly, we are not interested in elements that are already fixed in the population, so we condition on $0 < i < k$. Using conditional probability, realizing that if we ignore $i = 0$ and $i = k$, eq. 8 is actually $P[i,s,N,h \mid 0 < i < k]$, and using the above logic of Hardy–Weinberg sampling, we arrive at

$$P[i, s, N, h | 0 < i < k] = \frac{\binom{k}{i}\int_0^1 F[x](1 - (1-x)^2)^i((1-x)^2)^{k-i}dx}{1 - \int_0^1 F[x](1-x)^{2k}dx - \int_0^1 F[x](1 - (1-x)^2)^k dx}. \quad [9]$$

We applied these models to the TE-display data assuming $h = 0.5$ for all analyses (4), and using N_e values estimated from sequence data (5). The inbreeding statistic f was calculated, assuming the proportion of selfing was 0.001.

Demographic Modeling. Ross-Ibarra *et al.* (5) amplified and sequenced a set of 77 nuclear loci from 71 individual *A. lyrata* plants from the 5 populations sampled here and a 6th population from Iceland. Using this data, they built a pairwise demographic model (Fig. S3) in which each population was compared to Germany as a reference population. The model posits an ancestral population that evolves with a population mutation rate $\theta_A = 4N_{eA}\mu$, where N_{eA} is the ancestral effective population size and μ is the per nucleotide mutation rate. This ancestral population gives rise to two daughter populations at time τ_s generations in the past. The two daughter populations are initially bottlenecked and at size θ_{1b} and θ_{2b}, but recover from their bottlenecks to their modern population sizes of θ_1 and θ_2 at times τ_1 and τ_2 in the past. The population recombination rate ρ_A ($= 4N_{eA}c$) is assumed to be equal to the population mutation rate θ_A, and both μ and c, the per nucleotide recombination rate, are assumed to be invariable across populations (but vary among

loci). In total, the model consists of eight model parameters: the population mutation rate θ_1 of the reference Germany population, the ratios θ_2/θ_1, θ_A/θ_1, θ_{2b}/θ_1, and θ_{1b}/θ_1; the divergence time τ_s; and the bottleneck recovery times τ_1 and τ_2.

Ross-Ibarra *et al.* (5) used an approximate Bayesian approach to estimate the values of the parameters of the demographic model. Briefly, they simulated data under the proposed demographic model, drawing parameters values from specified prior distributions. They calculated summary statistics for each simulated data set (mean and variance of F_{ST}, shared segregating sites S_S, and the number of segregating sites S and nucleotide diversity π for each population) and compared these statistics with values from the observed data. Using a regression modification (6) of standard rejection sampling approaches, they then estimated posterior distributions for each parameter based on simulations with summary statistics most similar to the observed data.

1. Le QH, Bureau T (2004) Prediction and quality assessment of transposon insertion display data. *BioTechniques* 36, 222–228.
2. Petrov DA, Aminetzach YT, Davis JC, Bensason D, Hirsh AE (2003) Size matters: Non-LTR retrotransposoable elements and ectopic recombination in *Drosophila*. *Mol Biol Evol* 20:880–892.
3. Caballero A, Hill WG (1992) Effects of partial inbreeding on fixation rates and variation of mutant genes. *Genetics* 131:493–507.
4. Fry JD, Nuzhdin SV (2003) Dominance of mutations affecting viability in *Drosophila melanogaster*. *Genetics* 163:1357–1364.
5. Ross-Ibarra J, *et al.* (2008) Patterns of polymorphism and demographic history in natural populations of *Arabidopsis lyrata*. *PLoS ONE* 3:e2411.
6. Beaumont MA, Zhang W, Balding DJ (2002) Approximate Bayesian computation in population genetics. *Genetics* 162:2025–2035.

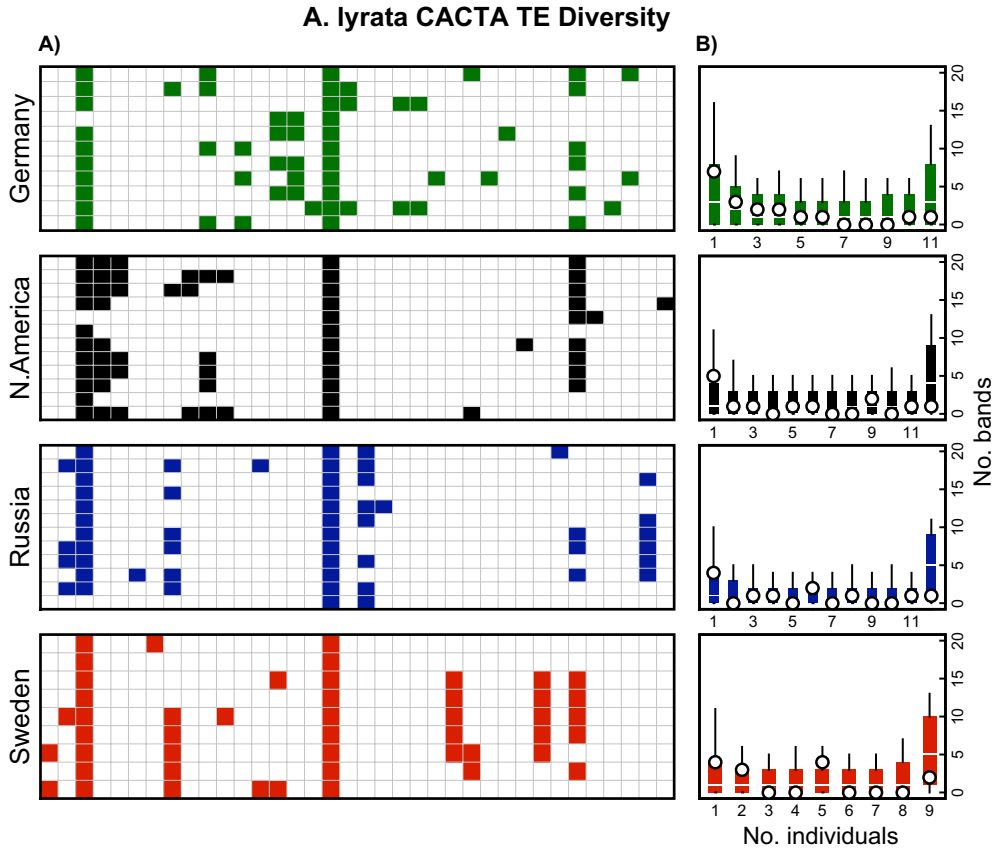

Fig. S1. Further examples of TE diversity data for five TE families. (*A*) Plots of TE-display data. A colored cell represents the presence of a TE, whereas a white cell represents a lack of TE detection. Each column represents a TE band, and each row represents an individual. Colors represent population of origin. (*B*) TE-band frequency spectra (BFS) for observed data (white circles) and simulated data (bars and vertical black lines). The bars represent the 95% credible interval, the white horizontal lines in the bars are the medians, and the black vertical lines show the full ranges of the simulations.

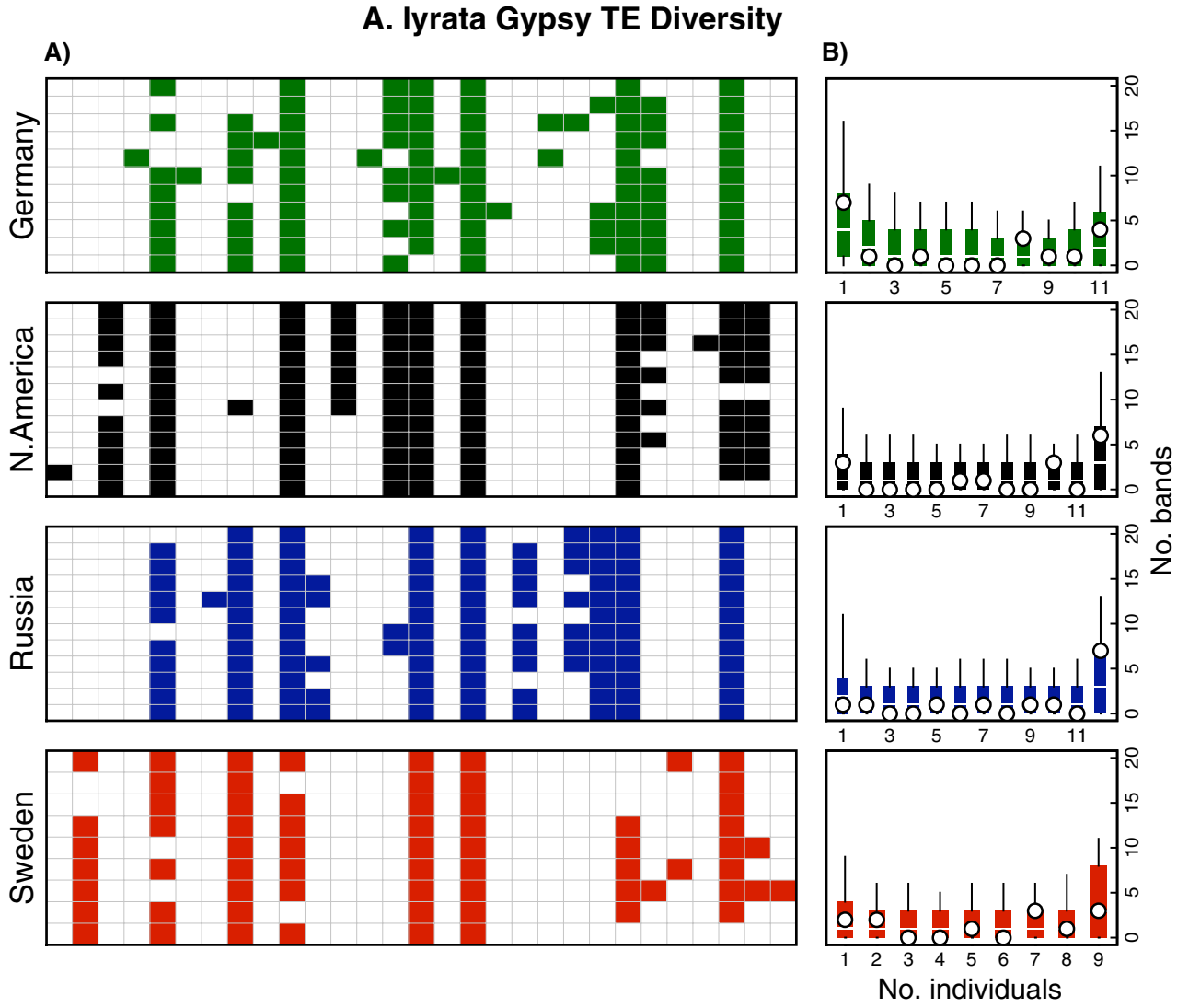

Fig. S1. (continued).

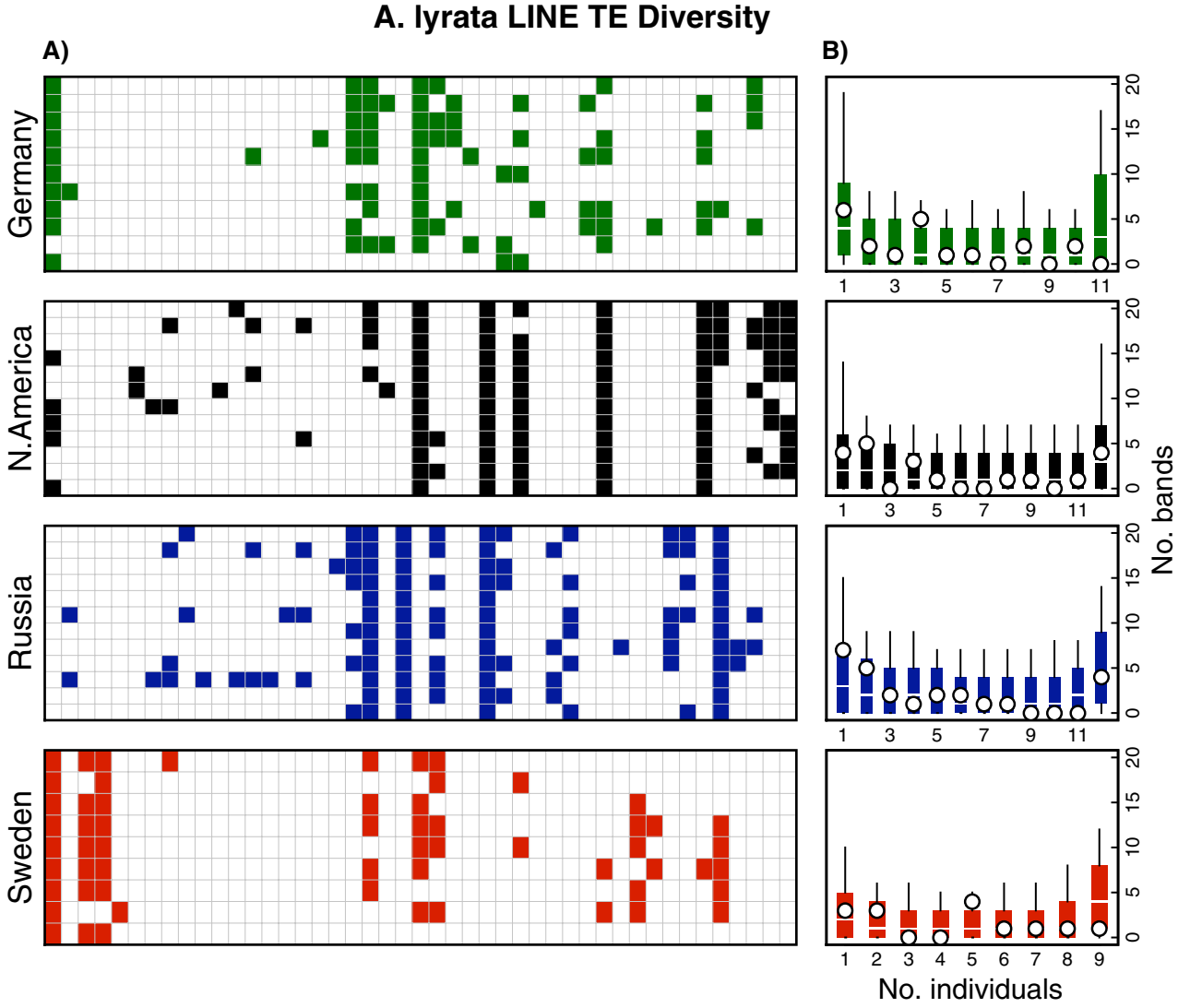

Fig. S1. (continued).

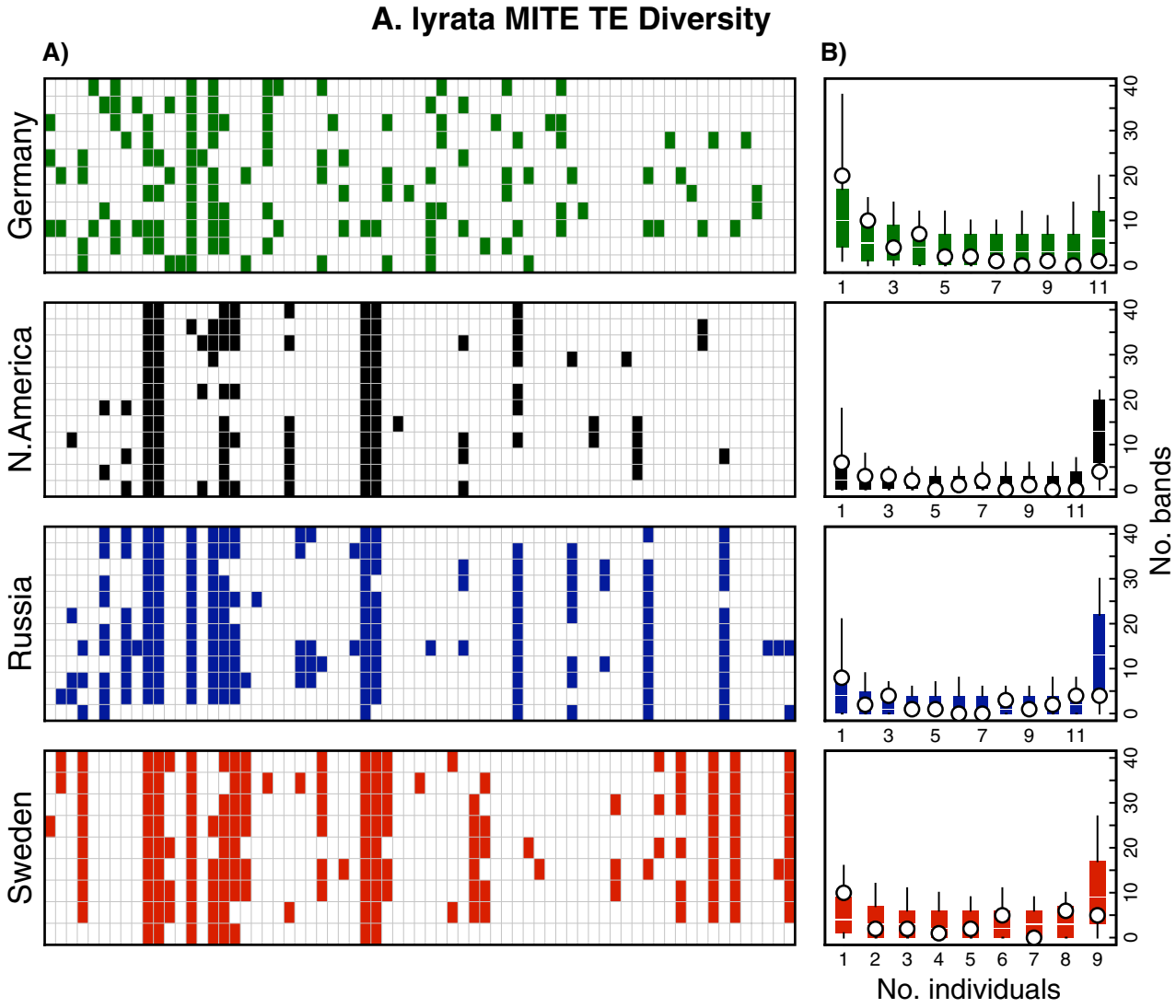

Fig. S1. (continued).

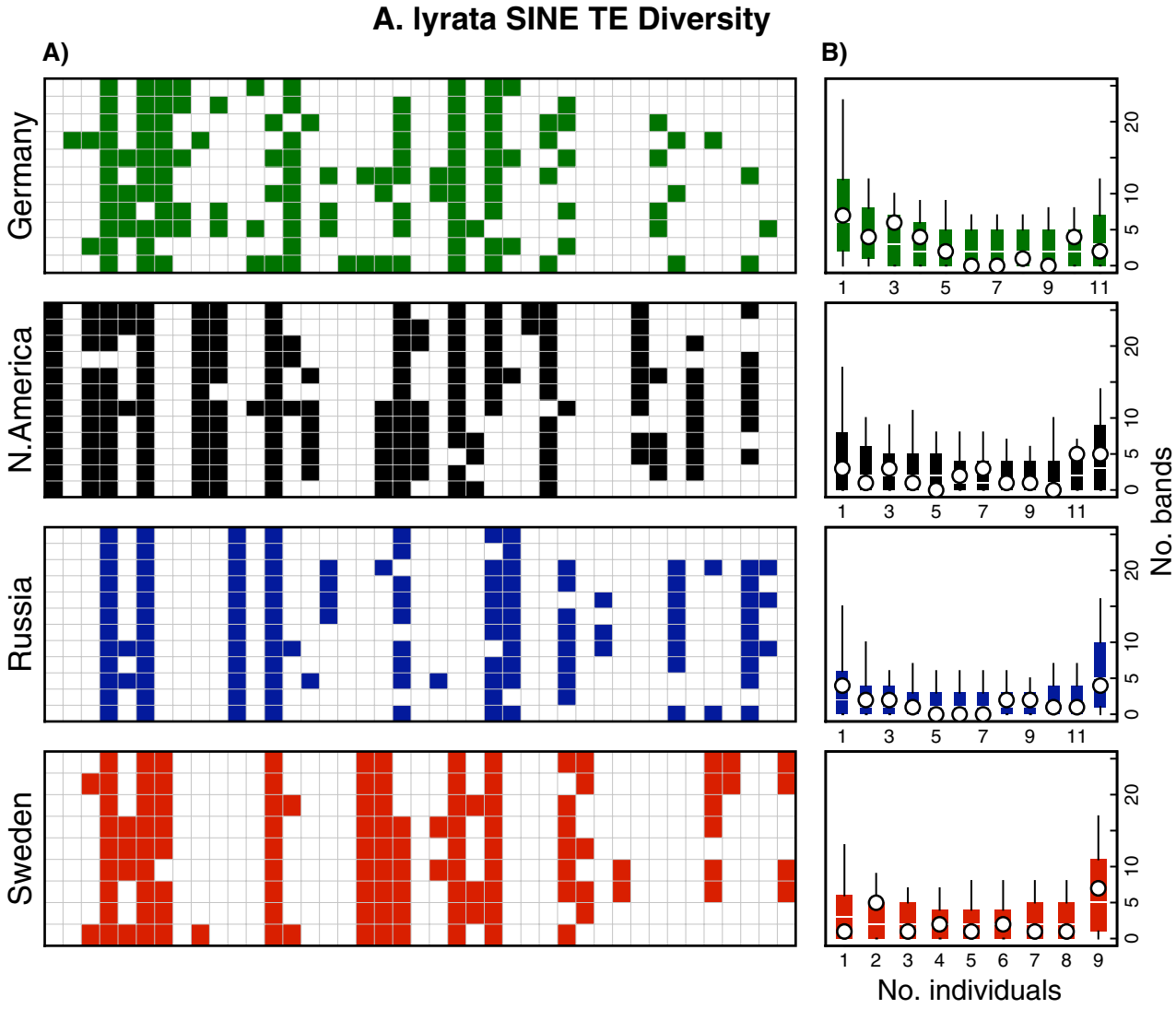

Fig. S1. (continued).

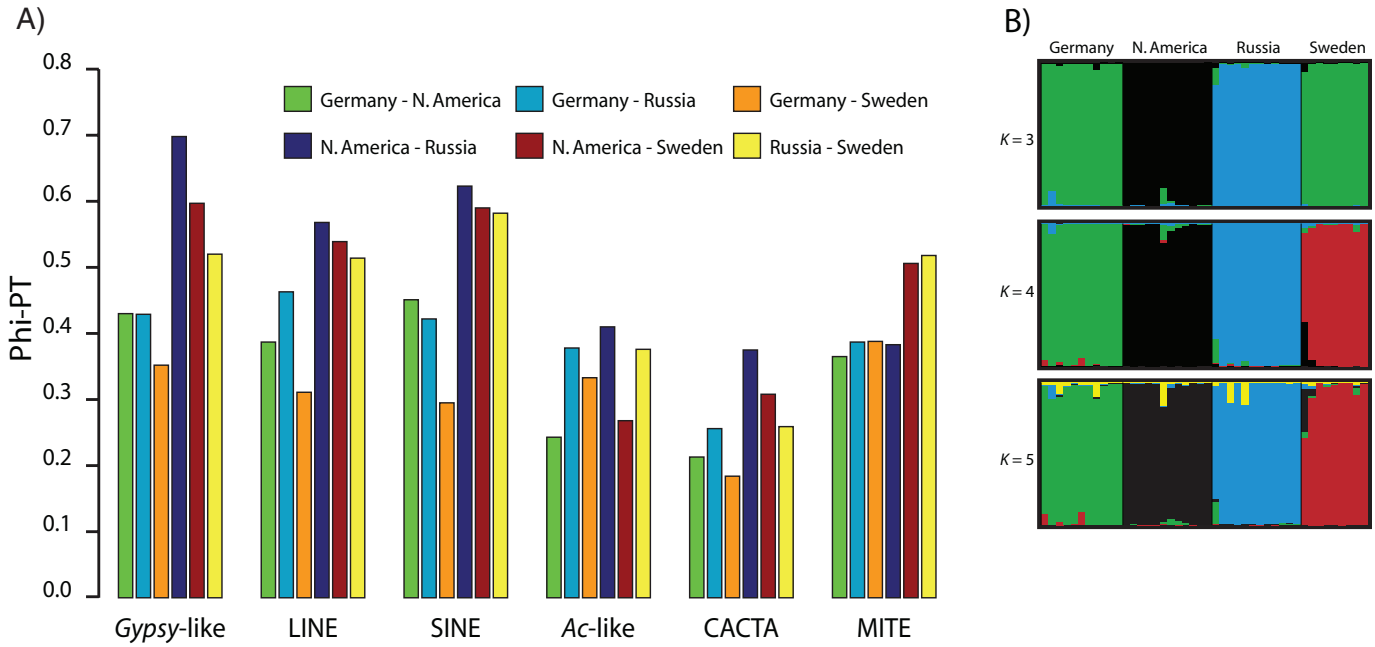

Fig. S2. (*A*) AMOVA Φ-PT measures for each pairwise population comparison in each TE family. (*B*) Results of a *Structure* analysis using all TE families as a single large unlinked dataset. Four clusters (*K*) yield the highest likelihood.

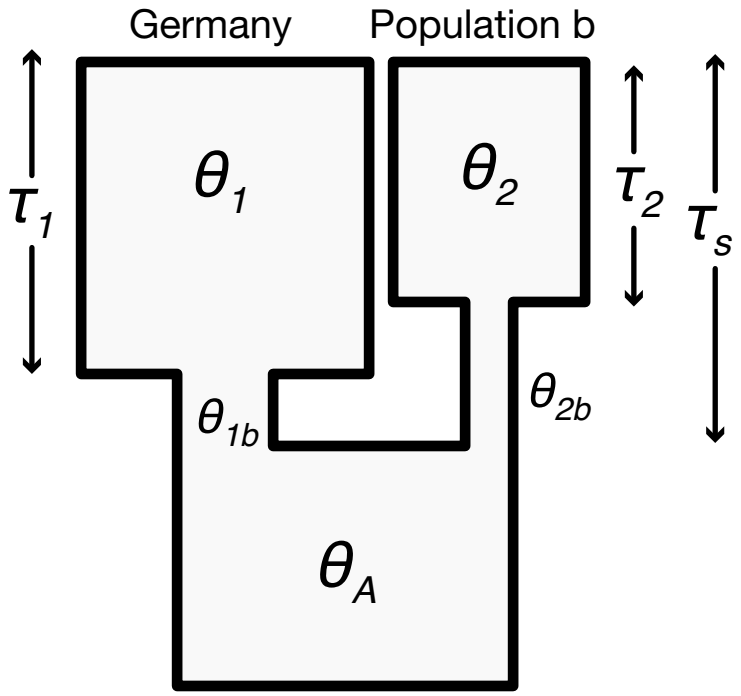

Fig. S3. Schematic representation of the bottleneck model used for parameter estimation.

Table S1. Maximum likelihood estimates of $N_e s$

TE family	$N_e \hat{s}$	95% C.I.
Germany		
Gypsy	0.68	−1.94, >5
LINE	0.27	−1.77, >5
SINE	0.07	−1.67, >5
Ac	−0.14	−1.96, 9.33
CACTA	−1.79	−4.27, 0.82
MITE	−2.60	−4.30, −1.19
North America		
Gypsy	4.75	−1.02, >5
LINE	−0.69	−2.83, >5
SINE	5.50	0.60, >5
Ac	−0.46	−1.88, 1.98
CACTA	0.21	−2.33, 5.50
MITE	−1.53	−3.81, 1.04
Russia		
Gypsy	2.70	−1.71, >5
LINE	−1.86	−4.10, 0.30
SINE	1.82	−1.24, >5
Ac	1.82	−0.70, >5
CACTA	0.14	−2.59, >5
MITE	6.00	1.04, >5
Sweden		
Gypsy	6.00	−0.88, >5
LINE	1.44	−1.54, >5
SINE	1.26	−1.51, >5
Ac	3.42	−0.25, >5
CACTA	−1.40	−4.56, >5
MITE	6.00	−0.06, >5

Table S2. Probabilities of rejecting null demographic model per TE site frequency

	Band frequency											
Population	1	2	3	4	5	6	7	8	9	10	11	12
Gypsy												
Ger	0.184	0.320	0.182	0.635	0.323	0.372	0.406	0.055	0.578	0.604	0.059	
N.Am	0.604	0.155	0.238	0.320	0.360	0.598	0.548	0.463	0.489	0.031	0.427	0.004
Rus	0.574	0.534	0.563	0.592	0.391	0.611	0.394	0.602	0.415	0.461	0.478	0.039
Swe	0.206	0.627	0.205	0.278	0.674	0.361	0.062	0.637	0.196			
LINE												
Ger	0.587	0.411	0.381	0.021	0.624	0.681	0.343	0.257	0.356	0.290	0.189	
N.Am	0.429	0.117	0.134	0.200	0.610	0.317	0.360	0.607	0.619	0.390	0.649	0.107
Rus	0.223	0.368	0.488	0.410	0.479	0.381	0.667	0.656	0.354	0.354	0.327	0.106
Swe	0.385	0.437	0.156	0.220	0.039	0.674	0.679	0.656	0.417			
SINE												
Ger	0.452	0.467	0.056	0.195	0.614	0.148	0.158	0.489	0.175	0.135	0.394	
N.Am	0.194	0.599	0.046	0.489	0.531	0.152	0.040	0.503	0.536	0.422	0.007	0.355
Rus	0.001	0.041	0.036	0.235	0.770	0.771	0.750	0.045	0.059	0.312	0.390	0.122
Swe	0.470	0.010	0.628	0.256	0.607	0.286	0.669	0.591	0.477			
Ac												
Ger	0.430	0.532	0.089	0.436	0.514	0.252	0.033	0.335	0.662	0.261	0.169	
N.Am	0.362	0.523	0.113	0.468	0.120	0.524	0.458	0.171	0.617	0.132	0.246	0.416
Rus	0.024	0.605	0.344	0.101	0.551	0.617	0.664	0.026	0.647	0.664	0.672	0.145
Swe	0.004	0.150	0.332	0.631	0.254	0.199	0.439	0.066	0.197			
CACTA												
Ger	0.236	0.454	0.507	0.359	0.668	0.594	0.416	0.448	0.456	0.569	0.632	
N.Am	0.418	0.334	0.509	0.264	0.672	0.624	0.426	0.464	0.183	0.473	0.563	0.581
Rus	0.425	0.096	0.520	0.634	0.338	0.233	0.424	0.529	0.487	0.478	0.555	0.596
Swe	0.311	0.528	0.150	0.226	0.027	0.352	0.370	0.365	0.455			
MITE												
Ger	0.037	0.131	0.495	0.086	0.375	0.463	0.275	0.087	0.305	0.071	0.106	
N.Am	0.229	0.457	0.368	0.567	0.171	0.463	0.527	0.172	0.449	0.147	0.103	0.244
Rus	0.288	0.370	0.297	0.315	0.355	0.127	0.136	0.315	0.407	0.624	0.012	0.029
Swe	0.219	0.145	0.264	0.153	0.423	0.145	0.060	0.072	0.574			

Observed data that fall outside of simulated 95% CI: $P < 0.025$. Bonferroni correction per TE family, $P = 0.05$ for 44 tests: $P < 0.001$.

Table S3. Comparisons of TE diversity to silent SNPs

	Number of polymorphic sites		Mean per diploid individual	
Population	TEs	Silent SNPs	TE bands	Silent SNPs*
Germany	157	571	53.72	354.53
N. America	129	151	58.67	130.93
Russia	117	169	57.24	126.15
Sweden	121	204	62.66	138.36

*Mean number of sites with one or more derived silent SNPs present per diploid individual.

Table S4. Oligonucleotide sequences

Name	Oligonucleotide
CAS1	5′-TAGCAAGGAGAGGACGCTGTCTGTCGAAGGTAAGGAACGGACGAGAGAAGGGAGA-3′
CAS2	5′-TCTTCCCTTCTCGAATCGTAACCGTTCGTACGAGAATCGCTGTCTCTCCTTGC-3′
AP1	5′-CGAATCGTAACCGTTCGTACGAGAATCGCT-3′
AP2 (FAM)*	5′-GTACGAGAATCGCTGTCTCTC-3′
AclII_1	5′-GGTTCGGTTAWTCGGTTAGCKG-3′
AclII_2	5′-GMTTCGGTTCGGTTAWTCGGTTAG-3′
CAC_1	5′-YTTTCGTAATGCTATGGTTGAAACACCTAAC-3′
CAC_2	5′-CATACAATTCTGACGCTATC-3′
GYPSY_1	5′-CGCAACAGAGACCCTCAA-3′
GYPSY_2	5′-CCCAGACTTAATCACCATTGA-3′
LINE_1	5′-AGGTGGAAGCTCTCCTCTGG-3′
LINE_2	5′-GTGTATGATTGGTGCCGACA-3′
SINE_1	5′-GTGACGCTTGGGACGAAA-3′
SINE_2	5′-CGAACCGGCGACTTCTAAGT-3′
MITE_1	5′-GCGGGAAAACGTGATTTT-3′
MITE_2	5′-CGCCAAAACCACAAAAATTA-3′

*AP2 labeled with a FAM fluorescent dye.