Population genomics and local adaptation in wild isolates of a model microbial eukaryote

Christopher E. Ellison[a], Charles Hall[a], David Kowbel[a], Juliet Welch[a], Rachel B. Brem[b], N. L. Glass[a], and John W. Taylor[a,1]

Departments of [a]Plant and Microbial Biology and [b]Molecular and Cell Biology, University of California, Berkeley, CA 94720-3102

Edited* by Jay C. Dunlap, Dartmouth Medical School, Hanover, NH, and approved January 7, 2011 (received for review October 7, 2010)

Elucidating the connection between genotype, phenotype, and adaptation in wild populations is fundamental to the study of evolutionary biology, yet it remains an elusive goal, particularly for microscopic taxa, which comprise the majority of life. Even for microbes that can be reliably found in the wild, defining the boundaries of their populations and discovering ecologically relevant phenotypes has proved extremely difficult. Here, we have circumvented these issues in the microbial eukaryote *Neurospora crassa* by using a "reverse-ecology" population genomic approach that is free of a priori assumptions about candidate adaptive alleles. We performed Illumina whole-transcriptome sequencing of 48 individuals to identify single nucleotide polymorphisms. From these data, we discovered two cryptic and recently diverged populations, one in the tropical Caribbean basin and the other endemic to subtropical Louisiana. We conducted high-resolution scans for chromosomal regions of extreme divergence between these populations and found two such genomic "islands." Through growth-rate assays, we found that the subtropical Louisiana population has a higher fitness at low temperature (10 °C) and that several of the genes within these distinct regions have functions related to the response to cold temperature. These results suggest the divergence islands may be the result of local adaptation to the 9 °C difference in average yearly minimum temperature between these two populations. Remarkably, another of the genes identified using this unbiased, whole-genome approach is the well-known circadian oscillator *frequency*, suggesting that the 2.4°–10.6° difference in latitude between the populations may be another important environmental parameter.

ecological genomics | genome scan | fungi | circadian clock

Discovering the genetic basis behind adaptive phenotypes has long been considered the holy grail of evolutionary genetics. Although there are now several studies that have succeeded in identifying genes responsible for such phenotypes, the majority of them use a "forward-ecology" approach whereby candidate genes are identified on the basis of their having a function related to conspicuous traits, such as pigmentation (1–4). A paucity of obvious phenotypic traits has been a major impediment for studying adaptation in microbes because these organisms are, by nature, inconspicuous. However, next-generation sequencing technology has made it possible for individual laboratories to acquire whole-genome sequence information across populations. This innovation has enabled an unbiased "reverse-ecology" approach whereby genes with functions related to ecologically relevant traits can be identified by examining patterns of genetic diversity within and between populations and identifying candidate genes as those showing the signature of recent positive selection and/or divergent adaptation between populations (5).

The feasibility of such an approach has been illustrated by several recent studies in macrobes including plants (6), insects (7, 8), mice (9), and fish (10). However, even in these cases, populations had been identified a priori on the basis of candidate phenotypes associated with tolerance for serpentine soil (6), assortative mating in nature (7, 8), the extremes of latitudinal clines (8), or morphology and geographic isolation (9, 10). By contrast, here we use comparative population genomics to simultaneously recognize populations de novo and identify candidate adaptive phenotypes.

We chose the filamentous, fungal genus *Neurospora* for this study because it is an ideal system for studying the evolutionary genomics of wild populations. Species within the genus are haploid, free-living heterotrophs with two sexes (*mat a* and *mat A*) and relatively small genomes (40 Mb) (11). Thousands of wild strains have been collected from around the world and are available from the Fungal Genetic Stock Center (FGSC), several phylogenies have been published that together provide broad taxon sampling across the genus (12–14), and there is a nearly complete gene deletion collection for *Neurospora crassa* (15).

Although *Neurospora* is a microbe, in terms of evolution, it is very similar to more developmentally complex animals. The genus is broadly distributed but also shows patterns of geographic endemism, and both intrinsic and extrinsic barriers to reproduction are acting to maintain species boundaries (12, 16). Additionally, Dettman et al. (17) have shown that reproductive isolation arises between strains of *Neurospora* evolved in the laboratory under different selective regimes, suggesting that local adaptation may be an important contributor to divergence between *Neurospora* populations in nature. Finally, unlike yeast, there is evidence that most species of *Neurospora* (including *N. crassa*) are highly outbred (18).

Here, we have discovered two previously unknown and recently diverged populations of *N. crassa* (Ascomycota) by resequencing transcriptomes from 48 individuals collected from the Caribbean basin. These two populations are exposed to different local environments (subtropical vs. tropical) and exhibit "islands" of divergence in genomic regions containing genes whose functions, patterns of nucleotide polymorphism, and null phenotype are consistent with local adaptation.

Results and Discussion

Population Genomics. We genotyped 48 isolates of *Neurospora crassa* from the Caribbean basin, South America, and Africa (Table S1) by identifying ≈135,000 SNPs from Illumina mRNA sequence tags. We estimated a SNP false-positive rate of 1/18,000 by sequencing mRNA from the reference strain (11). Using Bayesian clustering of allele frequencies (19) (Fig. S1) and phylogenetic inference using Bayesian methods (20) (Fig. 1), we found strong support for two cryptic populations in the dataset: one endemic to Louisiana and the other including isolates from Florida, Haiti, and the Yucatan (referred to as the Caribbean population). This genetic structure is also supported by our relatively high F_{ST} estimate of 0.19. These populations were not found by previous phylogenetic studies (12) and, in laboratory crosses, between-

Author contributions: R.B.B., N.L.G., and J.W.T. designed research; C.E.E., C.H., D.K., and J.W. performed research; C.E.E., C.H., D.K., R.B.B., and N.L.G. contributed new reagents/analytic tools; C.E.E. and C.H. analyzed data; and C.E.E., R.B.B., N.L.G., and J.W.T. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

Freely available online through the PNAS open access option.

Data deposition: Sequence Read Archive Accession SRA026962 and http://pmb.berkeley.edu/~taylor/ftp/Ellison_2011_SNPdata.txt.

[1]To whom correspondence should be addressed. E-mail: jtaylor@berkeley.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1014971108/-/DCSupplemental.

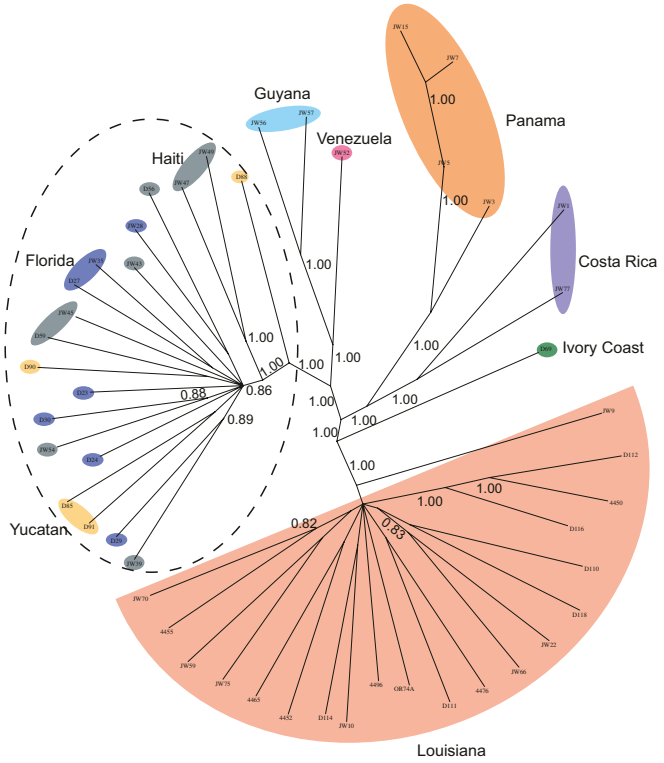

Fig. 1. Unrooted Bayesian phylogeny showing cryptic population structure. The tree was inferred from 135,035 polymorphic sites. The support values are clade posterior probabilities (converted to decimal) and values less than 0.75 are not shown. Each taxon is colored according to collection locality. The pink semicircle and dashed ellipse correspond to the Louisiana and Caribbean (Florida, Haiti, Yucatan) populations, respectively. The remaining clades likely represent additional populations that we are unable to resolve given their small sample size.

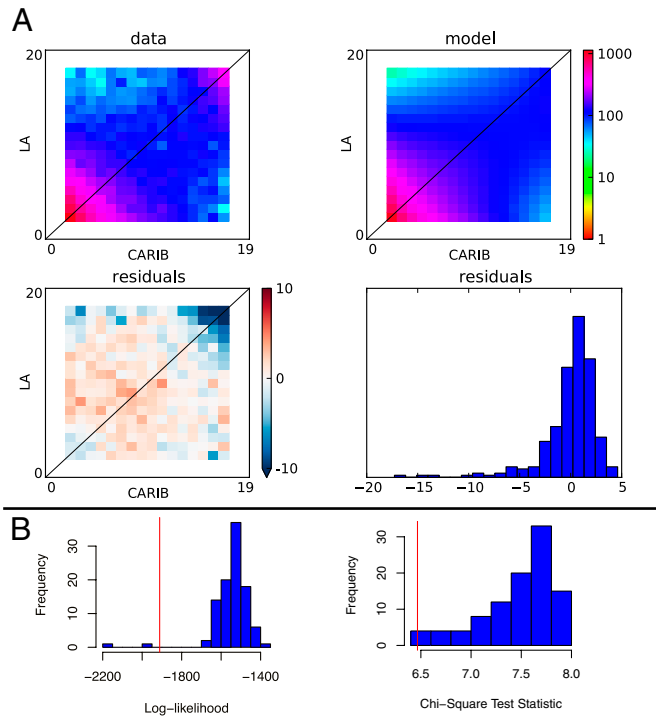

Fig. 2. Comparison of the *N. crassa* joint allele frequency spectrum with that expected under an isolation with migration demographic model. To determine whether our model of isolation with asymmetrical migration is appropriate for our data, we used two metrics to compare the joint allele frequency spectrum for the Louisiana and Caribbean populations of *Neurospora* with that expected under our model: (*A*) visual examination of the allele frequency spectra and calculation of the residuals between model and data, and (*B*) log-likelihood and Pearson's χ^2 goodness-of-fit tests. (*A*) Heat-maps showing the joint allele frequency spectrum for the Louisiana and Caribbean populations compared with that expected under a simple isolation with migration model. The color of a cell at position [X,Y] in the matrix corresponds to the number of derived alleles (relative to *N. tetrasperma*; *Materials and Methods*) that are at frequency X in the Caribbean population and frequency Y in the Louisiana population. The residuals represent the difference in the number of alleles predicted by the model compared with that found in the data for each bin in the spectrum (red, model predicts too many; blue, model predicts too few). The cluster of blue bins in the upper right corner of the heat-map of residuals implies that the model predicts too few high-frequency derived alleles, or in other words, there is an excess of high-frequency derived alleles in the data. (*B*) Goodness-of-fit tests using the likelihood and Pearson's χ^2 statistics and based on the results of 100 coalescent simulations in *ms* (22) under the demographic parameters inferred for the two *Neurospora* populations. Better fits have likelihood and χ^2 values closer to zero. The values from fitting the real data are shown by the red line.

population reproductive compatibility is indistinguishable from that within populations (16).

We used a diffusion-based approach implemented in the software package ∂a∂i (21) to infer demographic parameters for these two populations under an isolation with asymmetric migration model (Fig. 2 and Table 1). To assess the goodness of fit of this model and to obtain uncertainty estimates for the demographic parameters, we used the *Neurospora* parameters to simulate 100 datasets in *ms* (22). The optimized log-likelihood and the sum-of-squares of the residuals for the real data fall within the boundaries of those values from the simulated data, implying that our model is not grossly inappropriate for our data, nor is it an example of extreme overfitting (Fig. 2*B*). However, as shown in the heat-maps (Fig. 2*A*), there is an excess of high-frequency derived alleles in our data compared with what is predicted by the model. To determine whether this pattern was an artifact resulting from the misidentification of ancestral alleles, we fit our model to two additional datasets: in the first we included two outgroups (*Neurospora tetrasperma* and *Neurospora discreta*) (23, 24) and restricted our dataset to include only those SNPs for which both outgroups shared the same allele. In the second, we applied the misidentification correction that is part of the ∂a∂i package. We still observed an excess of high-frequency derived alleles in both of these cases, suggesting that this pattern is not an artifact (Fig. S2). A similar pattern has been found in wild populations of *Arabidopsis* and *Oryza* (25, 26), and Caicedo et al. (26) found that this pattern could be explained by either a complex demographic history including population bottlenecks and high migration rates or pervasive genome-wide positive selection in the form of selective sweeps, both of which are plausible scenarios for our *Neurospora* populations.

We infer a relatively high population migration rate from Louisiana into the Caribbean (0.77 effective migrants per generation) and approximately one sixth that rate in the other direction (Table 1). We also inferred a relatively recent divergence time (≈0.4 Mya) between the two populations, in agreement with the small proportion of fixed differences (9.4%; Table 2). Although we cannot eliminate the possibility that these populations diverged in complete allopatry, this scenario seems unlikely given the high dispersal potential of fungi (27) and the fact that the Louisiana population is closer to the Caribbean population (≈1,000 km) than some of the Caribbean localities are to each other (≈1,000–1,600 km). However, the strong support for population structure from multiple methods and the migration estimates from ∂a∂i both suggest that current migration is not sufficient to overcome genetic drift.

Table 1. Population demographic parameters

Parameter	Point estimate	Mean	SD
Ancestral N_e	344,049	343,700	9,822.7
LA N_e after split	1,151,961	1,211,000	488,347.4
CARIB N_e after split	364,841	387,000	209,200.4
Divergence time	494,547	514,500	165,472.5
Effective migration rate: CARIB into LA	0.1297	0.1639	0.1376
Effective migration rate: LA into CARIB	0.7651	0.9032	0.4705

Table shows the maximum-likelihood parameter estimates for an isolation with asymmetrical migration model fitted to the joint allele frequency spectrum for the Caribbean (CARIB) and Louisiana (LA) *N. crassa* populations. Uncertainty estimates were obtained by calculating the mean of the point estimate and the corresponding SD for each model parameter from 100 datasets simulated in *ms* (22) using the *Neurospora* maximum-likelihood parameters. N_e, effective population size.

Comparisons with *Saccharomyces*. This dataset also provides an important benchmark for comparison with the recent *Saccharomyces* population genomic study (28) (Table 3). Our population genetic summary statistics indicate that, as predicted previously, *Neurospora* is much more outbred than *Saccharomyces* (18). Linkage disequilibrium decays to half its maximum value at a physical distance of ≈0.78 kb, compared with ≈3 kb in *Saccharomyces cerevisiae* and ≈9 kb in *Saccharomyces paradoxus*. Additionally, nucleotide diversity within each *Neurospora* population is more than twofold greater than that found in the UK population of *S. paradoxus* and the Wine/European cluster of *S. cerevisiae*. We have also used the same approach as in Liti et al. (28) to estimate the number of deleterious nonsynonymous polymorphisms segregating in the *Neurospora* populations. Consistent with *Neurospora* being more outbred, we estimate that ≈34% of nonsynonymous polymorphisms are deleterious in the two *Neurospora* populations, which is approximately half the amount estimated in *Saccharomyces* (28).

Finally, ≈50% of the SNPs we identified are still segregating within each of the two *Neurospora* populations, whereas the majority of polymorphisms identified by Liti et al. are fixed within each nonmosaic *S. cerevisiae* lineage. Given that their effective population sizes are approximately an order of magnitude smaller than the estimates for *S. cerevisiae* (29) and *S. paradoxus* (30), the amount of ancestral variation that remains within these two *Neurospora* populations implies that they are much more recently diverged than any found in the *Saccharomyces* study. Such recent divergence makes these populations an ideal system for the study of incipient speciation and adaptation.

Genomic Islands of Divergence Between Populations of *Neurospora*. To identify candidate genomic islands of divergence, we conducted sliding window estimates of three different population genetic parameters: F_{ST} (31), Tajima's *D* (32), and Dxy (33). For each parameter, we identified empirical outliers in the 0.5% quantile. F_{ST} measures relative divergence and is the most commonly used metric in studies of heterogeneous genomic divergence (34). We additionally use Dxy, a measure of absolute divergence, following the recommendation of Noor and Bennet (35). To our knowledge, Tajima's *D* has not been previously used for this purpose but, in principle, scans of the combined populations should produce large positive values for regions showing low within-population polymorphism and high between-population divergence, making it similar to a relative divergence measure.

Table 2. Population genetic summary statistics

Populations	F_{ST}	$\theta\pi$	LD decay (kb)	D.S. (%)
Caribbean	NA	0.0023	0.85	34.0
Louisiana	NA	0.0024	0.70	33.1
Interpopulation	0.191	0.0029	NA	NA
S. paradoxus (U.K.)	NA	0.0010	≈9.0	NA
S. cerevisiae (W/E)	NA	0.0011	≈3.0	61.0

Summary statistics for the Caribbean and Louisiana populations of *N. crassa* compared with those for the Wine/European population of *S. cerevisiae* (W/E) and the UK population of *S. paradoxus* calculated by Liti et al. (28). $\theta\pi$ is the average number of pairwise differences between individuals; LD decay is the average physical distance over which the coefficient of linkage disequilibrium (r^2) decays to half its maximum value; and deleterious SNPs (D.S.) are an estimate of the number of amino acid changing polymorphisms that are deleterious, calculated as in Liti et al. (28). NA indicates that either the statistic is not applicable or the value was not calculated by Liti et al. (28).

Table 3. Summary of SNPs

Category	Percentage of total SNPs
Fixed in both populations	9.4
Polymorphic in both populations	48.7
Fixed in LA, polymorphic in CARIB	19.0
Fixed in CARIB, polymorphic in LA	22.9

We identified 135,035 SNPs from the pooled sequence information obtained from all isolates used in this study. This table shows the percentage of the total SNPs that fall into a given category. LA, Louisiana population; CARIB, Florida–Haiti–Yucatan population.

We were surprised to discover little overlap between the significant regions identified by the three different metrics. Out of a total of 37 regions, only two were identified by all three metrics (Fig. S3). Interestingly, removing from analysis the sites that fell within these regions still resulted in a phylogeny with strong support for the two populations (Fig. S4). As such, these two major loci are not the sole drivers for the population structure that we observe. This finding is consistent with the results of Bayesian clustering of allele frequencies (Fig. S1) whereby the two populations were delineated only according to differences in allele frequencies and lends credence to our model in which, despite the presence of gene flow between populations, genetic drift and/or natural selection has resulted in genome-wide differences in allele frequencies between populations.

Apart from these two candidate islands of divergence, we observed little overlap between the top-scoring regions across the three different metrics of population divergence applied to the data. A total of 35 regions were called significant in some but not all analyses. To investigate these discrepancies, we examined these regions in more detail.

Genomic loci that did not achieve consensus across our tests for divergence fit into three major classes: A, block-like haplotypes that do not perfectly sort by population; B, regions where relative divergence is high but absolute divergence does not stand out from the genomic background; and C, regions where absolute divergence is high but relative divergence does not stand out from the genomic background (Fig. S5). Patterns A and B were predicted by Noor and Bennett (35) and may result from an inversion or other barrier to recombination that was segregating in the ancestral population. In pattern A, both haplotypes are still segregating in each population. In pattern B, relative divergence is high because alternate haplotypes have become fixed in each population, but absolute divergence does not stand out from the genomic background because the region itself contains few polymorphic sites. Finally, the pattern described in C was only seen in Dxy outliers. Absolute divergence is high because there are a large number of polymorphisms in the region but relative divergence does not stand out from the genomic background because most of the polymorphisms are still segregating in both populations. We conclude that analyses of interpopulation divergence using only a single measure (e.g., F_{ST}) are susceptible to the identification of false positives. This result is troubling because it suggests that any such study, along with those using datasets of much lower resolution, will be unable to distinguish these misleading divergence outliers from true islands of divergence, potentially drastically overestimating the prevalence of this phenomenon (35).

Genes Inside Divergence Islands Have Functions and Patterns of Variation Consistent with Local Adaptation. The difference in latitude between the Louisiana and Caribbean populations suggests that they may have experienced differences in selective forces related to environmental parameters such as day length and average yearly minimum temperature [5.0 °C for Welsh, Louisiana and 13.8 °C for Homestead, Florida (36)]. We sought to investigate whether our candidate genomic islands of divergence between these two populations could harbor genetic factors that are locally adapted.

The first divergence island is on chromosome 3 and contains a pattern of nucleotide variation consistent with independent selective sweeps within each population: an excess of variants segregating at low frequency and reduced π (average number of intrapopulation pairwise differences) within both populations, relative to the flanking regions (Fig. 3 and Fig. S6). We find both the Caribbean and Louisiana haplotypes present among the outgroup strains (Fig. 3), but strains from the same locality always have the same haplotype. These facts are consistent with either a history of gene migration among populations or the presence of both haplotypes in the ancestral population, followed by the sweeping of a single haplotype to fixation within populations.

This region contains the genes *plc-1* (phospholipase C), an *MRH4*-like mitochondrial DEAD box RNA helicase, and the unnamed gene NCU06247 [inferred to encode an outer mitochondrial membrane protein (37)]. Coincidentally, Gavric et al. (38) observed this same pattern of divergence in *N. crassa plc-1*, but, lacking the context of the two different populations and the genome-wide sampling presented here, could not explain it. We also found another mitochondrial DEAD box RNA helicase (homolog of the yeast gene *MSS116*) as one of 12 genes in the Louisiana population that show the signature of positive selection by the McDonald-Kreitman (MK) test (39) (Table S2). We did not expect to find the *MRH4*-like RNA helicase in this case because the MK test is confounded by the reduced within-population polymorphism in the genomic islands of divergence.

RNA helicases are key factors in the microbial cold response (40), making it tempting to speculate that they are important to Louisiana *N. crassa*, which experience minimum temperatures almost 9 °C lower than their Caribbean relatives.

The second divergence island is on chromosome 7 and was identified by the highest observed values of all three divergence measures. It shows an unusually large number of variable sites, the majority of which are fixed between populations (Fig. 3). As in the chromosome three region, this pattern seemed to be consistent with the action of repeated selective sweeps within each population. This prediction holds true for the Louisiana population, in which Tajima's *D* for the region is negative and π decreases relative to the flanking regions (Fig. S7). This pattern, however, is not seen in the Caribbean Basin population (Fig. S7). Additionally, all non-Louisiana strains have the same haplotype, and the boundaries of the distinct region in the Louisiana population vary among individuals (Fig. 3). Together, these observations point to the introgression of a genomic region as a single "migrant tract" (41) into Louisiana from a more genetically diverged population or species that we did not sample. Under this model, the introgressed haplotype would have rapidly spread through the Louisiana population, explaining why nucleotide polymorphism within this region is reduced in this population but not the Caribbean, whereas the nonuniformity of the region's boundaries could be due to recombination that occurred after the introgression (41).

The sweep to fixation of this region within the Louisiana population implies that it contains a gene that may confer a local selective advantage over the ancestral haplotype. Among the five genes in this region is the circadian oscillator gene *frequency* (*frq*), the subject of a significant body of work using *N. crassa* as a model for understanding the circadian clock (e.g., refs. 42–44). Also present are an *NSL1*-like kinetochore MIND complex subunit, a *SEC14*-like phosphatidylinositol/phosphatidylcholine transfer protein, a *PAC10*-like prefoldin-α subunit, and a gene of unknown function (NCU02261). As with the helicases, it is tempting to speculate that *frq* is involved in adaptation, in this case related to differences in local photoperiod associated with the 2.4°–10.6° difference in latitude between the Louisiana population and various Caribbean population localities.

Characterizing the Candidate Adaptive Phenotypes. The distributions of these two populations in conjunction with the RNA helicase and major circadian oscillator that we find within these genomic islands of divergence suggest two major environmental factors that may be promoting local adaptation: temperature and day length. Here we have chosen to focus on the response to low temperature. We chose to focus only on low temperature, rather than both low and high temperature, for several reasons. The global distribution of *N. crassa* is mainly tropical, implying that the extension of its range into more temperate Louisiana is a derived condition (45). In addition, there is a 9 °C difference in the mean annual minimum temperature between Welsh, Louisiana and Homestead, Florida, but only a 0.7 °C difference in the mean annual maximum temperature (36). Thus, although winter

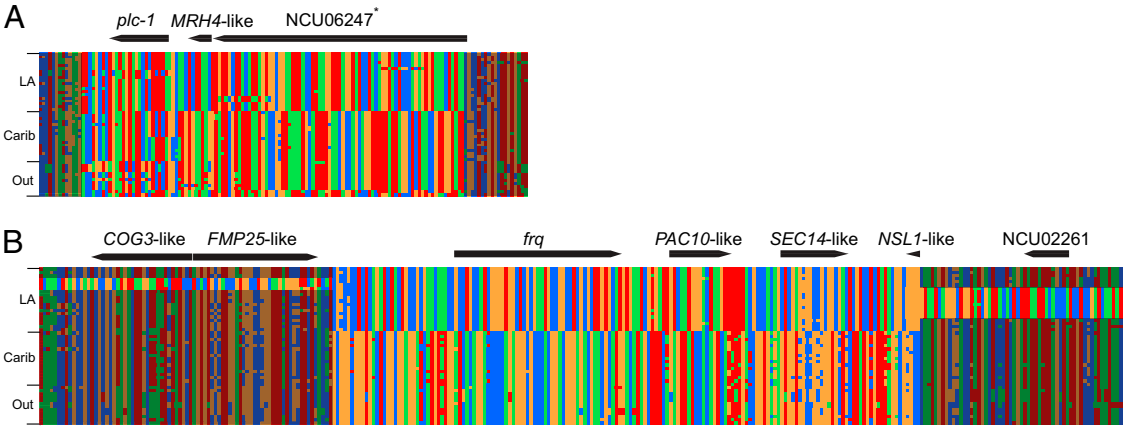

Fig. 3. Genomic islands of divergence genotype matrices. Each column is a polymorphic site, and each row contains the genotype for a particular strain. The flanking regions surrounding the divergence outliers are shaded and are shown to accentuate the distinct patterns of nucleotide polymorphism within the divergence outlier regions. Strains are grouped by population of origin. LA, Louisiana; Carib, Caribbean (Florida, Haiti, and the Yucatan); Out, outgroups from Central America, South America, and Africa. The matrix in *A* is the 10-kb divergence island on chromosome 3. The matrix in *B* is the 27-kb divergence island on chromosome 7. *The function of NCU06247 is unknown, but the protein localizes to the outer mitochondrial membrane (37).

in Louisiana is noticeably cooler than winter in the tropics, the summers are equally warm.

We predicted that individuals from the Louisiana population would exhibit higher fitness in cold temperature relative to individuals from the Caribbean. To test this prediction, we measured the growth rate of 10 randomly chosen individuals from the Louisiana population and 10 from the Caribbean population at 10 °C and 25 °C. For each individual, we calculated its growth rate at 10 °C as a percentage of its growth rate at 25 °C and found, as predicted, that the reduction in growth rate at 10 °C for strains from the Louisiana population is significantly less than that for strains from the Caribbean population, consistent with Louisiana strains exhibiting higher fitness at lower temperatures ($P = 0.031$; one-sided Mann-Whitney U test; Fig. 4*A*).

To begin to address the potential role in cold adaptation of the candidate genomic islands of divergence identified by our sequence analysis, we used strains from the *N. crassa* deletion collection (15) to determine whether genes in these islands were involved in low-temperature growth.

Preliminary growth experiments on null mutants of each locus at 10 °C suggested a cold temperature growth defect in deletions of the *MRH4*-like RNA helicase, the *PAC10*-like prefoldin subunit, and the unannotated gene NCU06247 (Fig. S8). To control for unlinked lesions introduced during generation of the deletion strains, and to verify reproducibility, we crossed each marked deletion strain to an unmarked tester strain and compared the growth rate of progeny with and without the deletion marker cassette. This experiment confirmed significant growth defects resulting from deletion of either of the two annotated genes but not NCU06247 (Fig. 4 *B–D*). The importance of the *MRH4*-like RNA helicase for growth at cold temperature in *N. crassa* is consistent with work on RNA helicases in many other systems (40, 46, 47), although the relatively modest effect of deleting this locus may be a consequence of functional redundancy among the 18 known helicases in the *N. crassa* genome, as has been suggested in *Arabidopsis* (47).

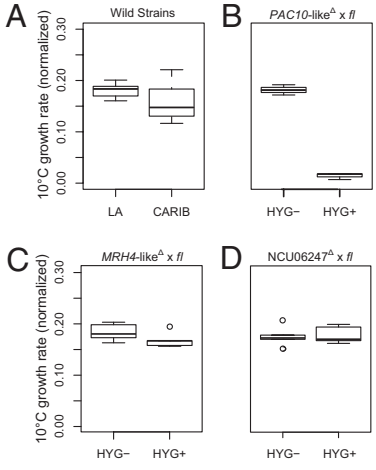

Fig. 4. Adaptation to cold temperature and functional characterization of genes within divergence islands. (*A*) For 10 strains from each population, growth rate at 10 °C was calculated as a percentage of that at 25 °C. Strains from Louisiana grow significantly faster than those from the Caribbean at 10 °C ($P = 0.031$; one-sided Mann-Whitney U test). (*B–D*) Three null mutant strains were crossed to the *fluffy* mating-type tester strain, and the growth rates of progeny with the hygromycin resistance deletion cassette were compared with progeny with the wild-type allele (hygromycin sensitive). The growth rate at 10 °C was calculated as a percentage of that at 25 °C. The progeny with the *PAC10*-like null allele and the *MRH4*-like null allele grew significantly slower at 10 °C, whereas those with the NCU06247 null allele did not (one-sided Mann-Whitney U test; $P = 0.05$, $P = 0.048$, and $P = 0.5$, respectively; sample size = 3 wild-type and 3 mutant progeny for *B*, 5 wild-type and 5 mutant progeny for *C* and *D*).

Taken together, our results indicate that the *MRH4*-like RNA helicase and the *PAC10*-like prefoldin subunit are critical for wild-type growth in cold temperatures in *N. crassa*, lending credence to the model that these genomic islands of divergence are the result of adaptation to low temperature. It should be noted, however, that large-scale fluctuations in climate have taken place since the divergence of these populations ≈0.4 Mya (48). Although there have been four major glacial events during this period, at ≈0.4 Mya the planet was in the middle of an interglacial period with an ice volume and surface temperature that is remarkably similar to current levels (48). In addition, although temperatures were cooler in absolute terms during the glaciations, paleontological studies based on pollen and plant microfossils indicate that the relative difference in temperature between the Florida peninsula and that of Louisiana was still present during the last glacial maximum (49). There is no evidence that this most recent glacial period was much more severe than those that preceded it (48), indicating that the Florida/Louisiana temperature difference likely was maintained since the divergence of these two populations.

Future work will be needed to establish the relationship between sequence variants at these loci, cold tolerance, and other environmental parameters, such as day length, that are relevant to the Caribbean and Louisiana populations. It is especially interesting to consider the possibility that the genes in these distinct genomic regions may be interacting in both the response to cold and the circadian rhythm given that the circadian clock of *N. crassa* exhibits temperature compensation and can be entrained by temperature in addition to light (50).

Conclusion

Here we have illustrated the utility of combining a "reverse ecology" genome-scan approach with functional characterization of the resulting candidate genes to identify ecologically relevant phenotypes in organisms that are difficult to study in nature. The major benefit of this approach, compared with a purely candidate gene approach, is that it provides a relatively unbiased look across the whole genome, allowing for identification of genes whose role in adaptation may not have been expected a priori. As it becomes easier to obtain large amounts of DNA sequence data, this type of approach is becoming increasingly common and will help facilitate the study of ecologically important nonmodel systems.

Although this approach has been demonstrated in other systems (6–10), here it has been used with a microbe, which is where it may prove to be the most useful. It can be difficult to apply this type of approach to populations of nonmodel organisms because it generally needs to be combined with a nearly complete reference genome assembly or an unfinished assembly paired with a genetic map (51). However, compared with macrobes, most microbes have smaller genomes with a lower repeat density, and low-cost, high-quality de novo genome assemblies from short reads have been achieved for both fungi (52) and bacteria (53). These features of microbes suggest that it is feasible to produce a reference genome assembly from a single individual while additionally resequencing many other individuals at low coverage to obtain polymorphism data that can be used for the genome scan. Furthermore, microbes are generally more amenable to genetic transformation, which may aid in the functional characterization of the candidate genes identified in the genome scan.

Materials and Methods

Identification of SNPs. Messenger RNA-Seq reads that did not map uniquely were discarded. Read alignments from each strain were pooled, and SNPs were identified using a Bayesian approach implemented in the program GigaBayes (54). To be included in the final set of high-quality SNPs, a candidate site was required to be biallelic and needed to meet or exceed the following criteria: coverage of five reads per allele, individual base qualities of 10, aggregate base qualities of 40, and Bayesian genotype probability of 0.90. To further reduce the number of potential false positives, singletons were discarded. Sites with missing data (i.e., the allele of one or more individuals was unidentifiable because it did not meet the above criteria) were excluded from analysis. Using these criteria, we found 5,640 genes that

had at least one SNP out of the ≈9,800 genes in the genome. These 5,640 genes had an average of 14.4 SNPs per gene.

Analysis of Population Demographics. Demographic parameters were estimated from the Louisiana and Caribbean joint allele frequency spectrum using a diffusion-based approach implemented in the program ∂a∂i (21). To control for the potential misidentification of ancestral states, we fit the model to two additional datasets: one in which we used two outgroups [*N. tetrasperma* FGSC #2508 (23) and *N. discreta* FGSC #8579 (24)] and one in which we applied a correction that is part of the ∂a∂i package. The results were nearly identical (Fig. S2), and we report the parameters estimated from the uncorrected spectrum.

Growth Rate Assays. All strains used in the growth rate assays were *a* mating type. The location of the hyphal front was recorded at regular intervals until it reached the other end of the tube. Each strain was grown in triplicate in constant darkness inside 25 °C and 10 °C incubators. Crosses involving null mutants were made to the *fluffy* mating type tester strain. The *fluffy* strain contains a mutation at a single locus that makes it aconidate and highly fertile (55). All progeny used in growth rate assays were screened to ensure that they produced macroconidia and thus did not have the *fluffy* mutation.

See *SI Materials and Methods* for more details.

ACKNOWLEDGMENTS. We thank M. Slatkin and D. Matute for helpful discussion and reading an earlier version of the manuscript; R. Gutenkunst for help with ∂a∂i; and the Fungal Genetic Stock Center for providing the strains used in this study. This work was supported by National Institutes of Health Grant NIGMS R01RGM081597 (to R.B.B., N.L.G, and J.W.T.) and the Chang-Lin Tien Graduate Fellowship (to C.E.E.).

1. Pool JE, Aquadro CF (2007) The genetic basis of adaptive pigmentation variation in *Drosophila melanogaster*. *Mol Ecol* 16:2844–2851.
2. Storz JF, et al. (2009) Evolutionary and functional insights into the mechanism underlying high-altitude adaptation of deer mouse hemoglobin. *Proc Natl Acad Sci USA* 106:14450–14455.
3. Nachman MW, Hoekstra HE, D'Agostino SL (2003) The genetic basis of adaptive melanism in pocket mice. *Proc Natl Acad Sci USA* 100:5268–5273.
4. Jiggins CD, McMillan WO (1997) The genetic basis of an adaptive radiation: Warning colour in two *Heliconius* species. *Proc R Soc Lond B Biol Sci* 264:1167–1175.
5. Li YF, Costello JC, Holloway AK, Hahn MW (2008) "Reverse ecology" and the power of population genomics. *Evolution* 62:2984–2994.
6. Turner TL, Bourne EC, Von Wettberg EJ, Hu TT, Nuzhdin SV (2010) Population resequencing reveals local adaptation of *Arabidopsis lyrata* to serpentine soils. *Nat Genet* 42:260–263.
7. Turner TL, Hahn MW, Nuzhdin SV (2005) Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biol* 3:e285.
8. Turner TL, Levine MT, Eckert ML, Begun DJ (2008) Genomic analysis of adaptive differentiation in *Drosophila melanogaster*. *Genetics* 179:455–473.
9. Harr B (2006) Genomic islands of differentiation between house mouse subspecies. *Genome Res* 16:730–737.
10. Hohenlohe PA, et al. (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genet* 6:e1000862.
11. Galagan JE, et al. (2003) The genome sequence of the filamentous fungus *Neurospora crassa*. *Nature* 422:859–868.
12. Dettman JR, Jacobson DJ, Taylor JW (2003) A multilocus genealogical approach to phylogenetic species recognition in the model eukaryote *Neurospora*. *Evolution* 57: 2703–2720.
13. Dettman JR, Jacobson DJ, Taylor JW (2006) Multilocus sequence data reveal extensive phylogenetic species diversity within the *Neurospora discreta* complex. *Mycologia* 98: 436–446.
14. Menkis A, Bastiaans E, Jacobson DJ, Johannesson H (2009) Phylogenetic and biological species diversity within the *Neurospora tetrasperma* complex. *J Evol Biol* 22:1923–1936.
15. Colot HV, et al. (2006) A high-throughput gene knockout procedure for *Neurospora* reveals functions for multiple transcription factors. *Proc Natl Acad Sci USA* 103: 10352–10357.
16. Dettman JR, Jacobson DJ, Turner E, Pringle A, Taylor JW (2003) Reproductive isolation and phylogenetic divergence in *Neurospora*: Comparing methods of species recognition in a model eukaryote. *Evolution* 57:2721–2741.
17. Dettman JR, Anderson JB, Kohn LM (2008) Divergent adaptation promotes reproductive isolation among experimental populations of the filamentous fungus *Neurospora*. *BMC Evol Biol* 8:35.
18. Powell AJ, Jacobson DJ, Salter L, Natvig DO (2003) Variation among natural isolates of *Neurospora* on small spatial scales. *Mycologia* 95:809–819.
19. Corander J, Marttinen P, Sirén J, Tang J (2008) Enhanced Bayesian modelling in BAPS software for learning genetic structures of populations. *BMC Bioinformatics* 9:539.
20. Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17:754–755.
21. Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD (2009) Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet* 5:e1000695.
22. Hudson RR (2002) Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics* 18:337–338.
23. Joint Genome Institute (2010) N. tetrasperma mat A v2.0. Available at http://genome.jgi-psf.org/Neute_matA2/Neute_matA2.home.html. Accessed March 10, 2010.
24. Joint Genome Institute (2010) N. discreta v1.0. Available at http://genome.jgi-psf.org/Neudi1/Neudi1.home.html. Accessed March 10, 2010.
25. Morton BR, Dar VU, Wright SI (2009) Analysis of site frequency spectra from *Arabidopsis* with context-dependent corrections for ancestral misinference. *Plant Physiol* 149:616–624.
26. Caicedo AL, et al. (2007) Genome-wide patterns of nucleotide polymorphism in domesticated rice. *PLoS Genet* 3:1745–1756.
27. Taylor JW, Turner E, Townsend JP, Dettman JR, Jacobson D (2006) Eukaryotic microbes, species recognition and the geographic limits of species: Examples from the kingdom Fungi. *Philos Trans R Soc Lond B Biol Sci* 361:1947–1963.
28. Liti G, et al. (2009) Population genomics of domestic and wild yeasts. *Nature* 458: 337–341.
29. Skelly DA, Ronald J, Connelly CF, Akey JM (2009) Population genomics of intron splicing in 38 *Saccharomyces cerevisiae* genome sequences. *Genome Biol Evol* 1: 466–478.
30. Tsai IJ, Bensasson D, Burt A, Koufopanou V (2008) Population genomics of the wild yeast *Saccharomyces paradoxus*: Quantifying the life cycle. *Proc Natl Acad Sci USA* 105:4957–4962.
31. Wright S (1950) Genetical structure of populations. *Nature* 166:247–249.
32. Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–595.
33. Nei M (1987) *Molecular Evolutionary Genetics* (Columbia Univ Press, New York).
34. Beaumont MA (2005) Adaptation and speciation: What can F(st) tell us? *Trends Ecol Evol* 20:435–440.
35. Noor MA, Bennett SM (2009) Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. *Heredity* 103:439–444.
36. Oregon Climate Service (2003) PRISM. Available at http://www.prismclimate.org. Accessed January 15, 2010.
37. Schmitt S, et al. (2006) Proteome analysis of mitochondrial outer membrane from *Neurospora crassa*. *Proteomics* 6:72–80.
38. Gavric O, dos Santos DB, Griffiths A (2007) Mutation and divergence of the phospholipase C gene in *Neurospora crassa*. *Fungal Genet Biol* 44:242–249.
39. McDonald JH, Kreitman M (1991) Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature* 351:652–654.
40. Hunger K, Beckering CL, Wiegeshoff F, Graumann PL, Marahiel MA (2006) Cold-induced putative DEAD box RNA helicases CshA and CshB are essential for cold adaptation and interact with cold shock protein B in *Bacillus subtilis*. *J Bacteriol* 188: 240–248.
41. Pool JE, Nielsen R (2009) Inference of historical changes in migration rate from the lengths of migrant tracts. *Genetics* 181:711–719.
42. McClung CR, Fox BA, Dunlap JC (1989) The *Neurospora* clock gene frequency shares a sequence element with the *Drosophila* clock gene period. *Nature* 339:558–562.
43. Aronson BD, Johnson KA, Loros JJ, Dunlap JC (1994) Negative feedback defining a circadian clock: Autoregulation of the clock gene frequency. *Science* 263:1578–1584.
44. Merrow M, Brunner M, Roenneberg T (1999) Assignment of circadian function for the *Neurospora* clock gene frequency. *Nature* 399:584–586.
45. Turner BC, Perkins DD, Fairfield A (2001) *Neurospora* from natural populations: A global study. *Fungal Genet Biol* 32:67–92.
46. Schade B, Jansen G, Whiteway M, Entian KD, Thomas DY (2004) Cold adaptation in budding yeast. *Mol Biol Cell* 15:5492–5502.
47. Kim JS, Kim KA, Oh TR, Park CM, Kang H (2008) Functional characterization of DEAD-box RNA helicases in *Arabidopsis thaliana* under abiotic stress conditions. *Plant Cell Physiol* 49:1563–1571.
48. Petit JR, et al. (1999) Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. *Nature* 399:429–436.
49. Jackson ST, et al. (2000) Vegetation and environment in Eastern North America during the Last Glacial Maximum. *Quat Sci Rev* 19:489–508.
50. Liu Y, Bell-Pedersen D (2006) Circadian rhythms in *Neurospora crassa* and other filamentous fungi. *Eukaryot Cell* 5:1184–1193.
51. Storz JF, Wheat CW (2010) Integrating evolutionary and functional approaches to infer adaptation at specific loci. *Evolution* 64:2489–2509.
52. Nowrousian M, et al. (2010) De novo assembly of a 40 Mb eukaryotic genome from short sequence reads: *Sordaria macrospora*, a model organism for fungal morphogenesis. *PLoS Genet* 6:e1000891.
53. Reinhardt JA, et al. (2009) De novo assembly using low-coverage short read sequence data from the rice pathogen *Pseudomonas syringae pv. oryzae*. *Genome Res* 19: 294–305.
54. Hillier LW, et al. (2008) Whole-genome sequencing and variant discovery in *C. elegans*. *Nat Methods* 5:183–188.
55. Lindegren CC (1933) The genetics of *Neurospora* III: Pure bred stocks and crossing over in *N. crassa*. *Bull Torrey Bot Club* 60:133–154.