ARTICLE

Signatures of Purifying and Local Positive Selection in Human miRNAs

Hélène Quach,[1] Luis B. Barreiro,[1,6] Guillaume Laval,[1] Nora Zidane,[2] Etienne Patin,[1] Kenneth K. Kidd,[3] Judith R. Kidd,[3] Christiane Bouchier,[2] Michel Veuille,[4] Christophe Antoniewski,[5] and Lluís Quintana-Murci[1,*]

MicroRNAs (miRNAs) are noncoding RNAs involved in posttranscriptional gene repression, and their role in diverse physiological processes is increasingly recognized. Yet, few efforts have been devoted to evolutionary studies of human miRNAs. Knowledge about the way in which natural selection has targeted miRNAs should provide insight into their functional relevance as well as their mechanisms of action. Here we used miRNAs as a model system for investigating the influence of natural selection on gene regulation by characterizing the full spectrum of naturally occurring sequence variation of 117 human miRNAs from different populations worldwide. We found that purifying selection has globally constrained the diversity of miRNA-containing regions and has strongly targeted the mature miRNA. This observation emphasizes that mutations in these molecules are likely to be deleterious, and therefore they can have severe phenotypic consequences on human health. More importantly, we obtained evidence of population-specific events of positive selection acting on a number of miRNA-containing regions. Notably, our analysis revealed that positive selection has targeted a "small-RNA-rich island" on chromosome 14, harboring both miRNAs and small nucleolar RNAs, in Europeans and East Asians. These observations support the notion that the tuning of gene expression contributes to the processes by which populations adapt to specific environments. These findings will fuel future investigations exploring how genetic and functional variation of miRNAs under selection affects the repression of their mRNA targets, increasing our understanding of the role of gene regulation in population adaptation and human disease.

Introduction

MicroRNAs (miRNAs) are a class of small noncoding RNA molecules that have revealed a new dimension in the complexity of regulation of gene expression in eukaryotes.[1–6] miRNAs are ~22 nucleotide (nt) single-stranded RNAs that generally silence gene expression by binding to target messenger RNAs (mRNAs). miRNA-mRNA binding usually involves strong base pairing between the 5′ end of a miRNA and its target complementary sequence in the 3′-untranslated region (3′-UTR) of an mRNA. miRNA binding appears to result in translational repression and, in some cases, degradation of cognate mRNAs, thus disrupting, fully or partially, the expression of the respective protein-coding genes. Production of animal miRNAs is a two-step process where, first, cleavage of the primary miRNA transcript in the nucleus by the RNase III Drosha liberates a short stem-loop RNA called miRNA precursor (pre-miRNA).[7,8] After nuclear export by Exportin-5, the typical stem-loop structure is recognized and accurately processed by specialized endonucleases of the Dicer protein family to yield a discrete ~22 nucleotide duplex termed the miRNA/miRNA* duplex.[8–11] One of these two strands, the mature miRNA (or miR), is then coupled to a second endonuclease in the Argonaute family of proteins.[12,13] Ultimately, the mature miRNA serves as a guide to direct Argonaute proteins to target mRNAs based upon Watson-Crick base pairing between the miRNA and its target.

Increasing evidence suggests that miRNAs regulate a wide range of biological processes and human diseases.[14] They have been implicated in cell growth, tissue differentiation, cell proliferation, embryonic development, insulin secretion, immune response, and apoptosis.[1,3,14,15] A strong link between miRNAs and cancer has been demonstrated more recently: miRNAs may also act as tumor suppressor genes and oncogenes.[16–18] The recognition of the importance of miRNA-mediated gene regulation in such diverse physiological processes suggests that miRNA-related genetic alterations underlie more human diseases than are currently appreciated. Today, there have been reported ~70 diseases associated with miRNAs, and these range from Parkinson disease to cancer and heart failure.[19–21] All of these observations have led to an explosion of studies aiming to identify new miRNAs in humans and to provide better understanding of their involvement in human disease. Most computational analyses have suggested that there are more than 1,000 miRNAs in humans.[22–24] To date, 678 human miRNAs, which are listed in the official miRNA database (miRBase),[25] have been characterized by various techniques, including small RNA cloning and deep-sequencing-based approaches.

[1]Institut Pasteur, Human Evolutionary Genetics, Centre National pour la Recherche Scientifique, URA3012, F-75015, Paris, France; [2]Institut Pasteur, Genomics Platform, F-75015, Paris, France; [3]Department of Genetics, Yale University School of Medicine, New Haven, CT 06510, USA; [4]Ecole Pratique des Hautes Etudes, UMR 5202, Equipe de génétique des populations, Muséum National d'Histoire Naturelle, F-75005, Paris, France; [5]Institut Pasteur, Genetics and Epigenetics of Drosophila, Centre National pour la Recherche Scientifique, URA 2578, F-75015, Paris, France
[6]Present address: Department of Human Genetics, University of Chicago, Chicago, IL 60637, USA
*Correspondence: quintana@pasteur.fr
DOI 10.1016/j.ajhg.2009.01.022. ©2009 by The American Society of Human Genetics. All rights reserved.

Computational methods have estimated that approximately 20% to 30% of all human genes are targets of miRNA regulation, and an average of 200 targets per miRNA has been predicted.[2,26,27]

Assessing the levels of naturally occurring variation in miRNAs and understanding the way in which natural selection has targeted them should provide insight into both the gene-regulation mechanisms mediated by these molecules and their functional relevance. Likewise, it should help in identifying functionally important variants potentially associated with disease. In contrast with the effort devoted to the identification, characterization, and functional analysis of these molecules, evolutionary and population genetics data on human miRNAs remain scarce. Cross-species sequence comparisons of 122 miRNAs from ten primate species have indicated the high level of conservation of miRNA genes and a striking decrease of conservation for sequences flanking the miRNA hairpins.[23] Although the strong conservation observed across primate miRNAs suggests that natural selection has selected against most mutations at these molecules, selective pressures are not static over time and can vary in a species-specific manner. New miRNA mutations can be advantageous in humans and reach high frequencies through the action of positive selection. Alternatively, they can be neutral (or quasi-neutral) and increase in frequency in the population by simple genetic drift. For example, a recently identified common G/C polymorphism (rs2910164) within the sequence of the human pre-miR-146a (*MIR146A*), which is highly conserved throughout species, reduces the amount of mature miR-146a and contributes to the genetic predisposition to papillary thyroid carcinoma (PTC [MIM 188550]).[28]

In humans, few attempts have been made to describe the levels of naturally occurring genetic variation in miRNAs. A bioinformatics screening of the dbSNP database has shown that the density of single-nucleotide polymorphisms (SNPs) is lower in human miRNA loci with respect to their flanking regions.[29] However, the analysis of SNP public databases presents some limitations related to the nature of genotyping data: (i) the caveat of SNP ascertainment bias mainly due to an underrepresentation of low-frequency variants and (ii) no full coverage of the whole extent of sequence variation. By contrast, sequence-based methods are unbiased in that they allow the characterization of the full extent of genetic variation (e.g., low-frequency variants) and, therefore, the performance of most statistical tests designed to detect natural selection. Because no systematic resequencing effort of human miRNAs from diverse populations worldwide has been performed so far, the evolutionary forces driving human miRNA diversity are still poorly described. In addition, no attempt has been made to elucidate whether positive selection has targeted human miRNAs in a population-specific manner.

Here, we investigated the extent to which natural selection has acted upon miRNAs by characterizing the full spectrum of naturally occurring sequence variation in 117 human miRNAs from different populations, including sub-Saharan Africans, Europeans, and East Asians. Our resequencing data allowed us to determine the global levels of polymorphisms in miRNA-containing regions and to shed light on the selective pressures targeting miRNAs during recent human evolution. Our analyses provided evidence that natural selection, in its different modes, has affected the evolution of human miRNAs. Specifically, purifying selection has constrained variation in miRNAs—particularly in mature miRNAs—at the level of the entire human species, highlighting the essential role of these molecules in modulating gene-regulatory networks. In addition, we showed that population-specific positive selection has targeted some miRNA-containing regions and has contributed to the adaptation of different human populations to their idiosyncratic environments.

Material and Methods

DNA Samples

Sequence variation for the 117 miRNAs was surveyed in a total of 91 individuals from populations chosen from different geographical regions. The description of the specific population samples is available from the ALFRED website. From sub-Saharan Africa, we chose Yoruba from Nigeria (24 individuals from UID "SA000036J") and Chagga from Tanzania (24 individuals from UID "SA000487T"), from Europe we chose Danes (23 individuals from UID "SA000046K"), and from East Asia we chose Han Chinese (20 individuals from UID "SA000009J"). All individuals were healthy donors who gave informed consent. Orthologous miRNA regions were also sequenced from one chimpanzee (*Pan troglodytes*) when the corresponding sequences were not publicly available. This study was approved by the Institut Pasteur Institutional Review Board (n° RBM 2008.06).

Data Generation

Resequencing of miRNA-Containing Regions

We resequenced 92 miRNA-containing regions for each individual; these encompassed 117 different miRNA hairpins and constituted 110 kb of diploid sequence in total per individual (Table S1 in the Supplemental Data). Amplifications were performed in 20 μl PCR reactions with *TaKaRa LA Taq*. All the PCR products (~1.2 kb) were sequenced with the BigDye Terminator v1.1 Cycle Sequencing Kit and the 3730*xl* DNA Analyzer from Applied Biosystems. Detailed information on primers and PCR conditions is available on request. Sequence files and chromatograms were analyzed with GENALYS software.[30] For quality control, and to avoid allele-specific amplification, we designed new primers and repeated sequence reactions when new mutations were identified in primer-binding regions. All singletons or ambiguous polymorphisms were systematically reamplified and resequenced. The 117 miRNA hairpin structures were chosen from the miRBase database (release 8.0). miRNAs were chosen independently of their functions, but preference was given to miRNAs located in genomic regions allowing the amplification of a single miRNA within a ~1.2 kb PCR fragment. In total, 77 of the 92 regions sequenced contained a single pre-miRNA, whereas 15 regions contained multiple pre-miRNAs: there were 11 with two pre-miRNAs each, two with three

pre-miRNAs each, and two others with six pre-miRNAs each. Details of miRNA nomenclature, states, and chromosomal location for each miRNA are provided in Table S1.

Resequencing of Independent Autosomal Noncoding Regions

Twenty autosomal noncoding regions were amplified and sequenced in the same population panel used for the sequencing of miRNA-containing regions. The mean size of each of these regions was 1.3 kb, accounting for an additional 26.4 kb of diploid sequence per individual. The genomic localization of these regions can be found in Table S2. The 20 autosomal noncoding regions are dispersed throughout the genome and were chosen (i) to be independent from each other, (ii) to be at least 200 kb away from any known gene, predicted gene, or spliced expressed sequenced tag (EST), and (iii) not to be in linkage disequilibrium (LD) with any known gene or spliced EST. In this view, it is unlikely that these regions have been targeted by natural selection (even through hitchhiking), and therefore, their patterns of diversity should only mimic the demographic history (i.e., genetic drift) of the populations under study.

Statistical Analyses

Population-Genetics Statistics Used for Summarizing miRNA Sequence Diversity

Haplotypes were reconstructed from unphased genotypic data, using the Bayesian method implemented in PHASE software.[31] Summary statistics, such as the number of segregating sites (S), haplotype diversity (H), the nucleotide diversity per site π, and θ_W were obtained with DnaSP.[32] SNP density per region was calculated as the ratio between the number of segregating sites observed and the sequence length of the region. Sequence-based neutrality tests (i.e., Tajima's D, Fu and Li's F^*, and Fay and Wu's H) and the sliding-window analysis of nucleotide diversity (π) were also performed with DnaSP.[32] Population differentiation (F_{ST}) for each of the miRNA-containing regions was assessed with Arlequin.[33] We used the iHS statistic[34] to search for recent positive-selection events. We obtained iHS values for each candidate SNP by using Haplotter.[34] For the small-RNA-rich island, iHS values were obtained for 157, 156, and 147 SNPs, for Africans, Europeans and East Asians, respectively.

Assessment of Statistical Significance

To assess whether the mean nucleotide diversity (π) at miRNA-containing regions was significantly different from that observed at the 20 noncoding regions (Tables S3 and S4), we performed the nonparametric Mann-Whitney test both for all populations together and for individual population groups. To assess whether genetic diversity was significantly lower in the miRNA hairpins with respect to flanking regions, we compared the mean of nucleotide diversity π obtained from 87 nt sliding windows, which is equal in size to the mean length of the studied miRNA hairpins (i.e., the pre-miRNA and the basal stem, Figures 1A and 1B), to a distribution of randomly resampled values. These resampled values were obtained as follows: for each of the 77 sequenced regions containing a single miRNA, an 87 nt fragment was randomly and independently sampled; for each of the 15 sequenced regions containing multiple miRNAs, we resampled as many 87 nt fragments as miRNAs contained in each region (e.g., for a sequenced region containing three miRNAs, we resampled three fragments of 87 nt each). We calculated π for each fragment and then averaged over all 117 resampled fragments. We resampled sets of 117 values 10,000 times, so that we obtained a distribution of 10,000 values of the mean π calculated on 117 randomly resampled 87 nt fragments. We then estimated p values by comparing the 87 nt sliding-window values with these distributions. To test for a significant decrease in mean SNP density and nucleotide diversity in the various domains within the 87 nt miRNA hairpins (Figures 1C and 1D), we followed the same resampling procedure described above. We thus compared the mean SNP density and nucleotide diversity for the various hairpin domains to a distribution of randomly resampled values. Fragments of 32, 22, and 11 nt were resampled for the stem, miR/miR* and loop domains, respectively (Figure 1A). Of note, the mean SNP density and nucleotide diversity were calculated for the actual sizes of each of the different domains within the miRNA hairpins, whereas the sizes of the resampled fragments (i.e., 32, 22 and 11 nt) correspond to the mean sizes observed in our dataset for each of the different hairpin domains. We assessed the significance of the complete lack of diversity observed in the first 14 nt of the miR by using the same resampling procedure and considering a resampled fragment of 14 nt.

Coalescent Simulations of Demographic Models

p values for the various neutrality tests and population differentiation index F_{ST} were estimated from coalescent simulations under a finite-site neutral model of evolution with no recombination, which has been shown to be a conservative assumption.[35] Coalescent simulations were performed with SIMCOAL.[36] Each coalescent simulation was conditional on the observed sample sizes (48, 23, and 20 individuals sampled in Africa, Europe, and East Asia, respectively). We simulated the ~1,200 bp of each miRNA-containing region by modeling a per -site mutation rate (μ) across loci by sampling μ from a Gamma distribution with a mean of 2.0×10^{-8} and a 95% confidence interval of 1.5×10^{-8} to 4.0×10^{-8} (Voight et al.[37]). We first performed a set of 10^4 simulations under the assumption that populations had constant effective sizes that were gamma distributed[37] and that the mean effective population size was 10,000 individuals (95% CI: 3,000–21,000). We computed the p values by counting the number of simulated values of the different summary statistics that do not fall into the range of observed values (Table S5).

We next assessed the impact of demographic history on the robustness-of-neutrality tests and population-differentiation index F_{ST}. We simulated data under a demographic model inferred from the diversity patterns for the twenty noncoding regions (Tables S4 and S6) by using an approximate Bayesian computation (ABC) approach[38–40] that estimates the demographic parameters that best explain the patterns of diversity observed at the noncoding regions. This demographic model considered an out-of-Africa exit of modern humans through a single major dispersal event occurring (T_{OoA}) 60,000 years before present (YBP) (95% CI: 47,000–85,000 YBP), a bottleneck consisting of a decrease in population size (β_{OoA}) by a factor of 5 (95% CI: 2.6 - 8.8) after the out-of-Africa exodus for non-Africans, and an expansion starting (t_A) 27,000 YBP (95% CI: 20,000–40,000 YBP) from an ancestral population (N') of 13,800 individuals (95% CI: 9,000–19,800) with a per-generation exponential expansion rate (α_A) of 0.007 (95% CI: 0.002–0.016) for Africans (Figure S1). In addition, this demographic model considered that the ancestors of modern Europeans and East Asians diverged at ($T_{E\text{-}EA}$) 22,500 YBP (95% CI: 17,500–35,000 YBP) from the population of ancestral migrants (Figure S1). We computed the p values of the different neutrality tests for the miRNA regions corrected by this demographic model by counting the number of simulated values of the different summary statistics that did not fall into the range of observed values (Table S5).

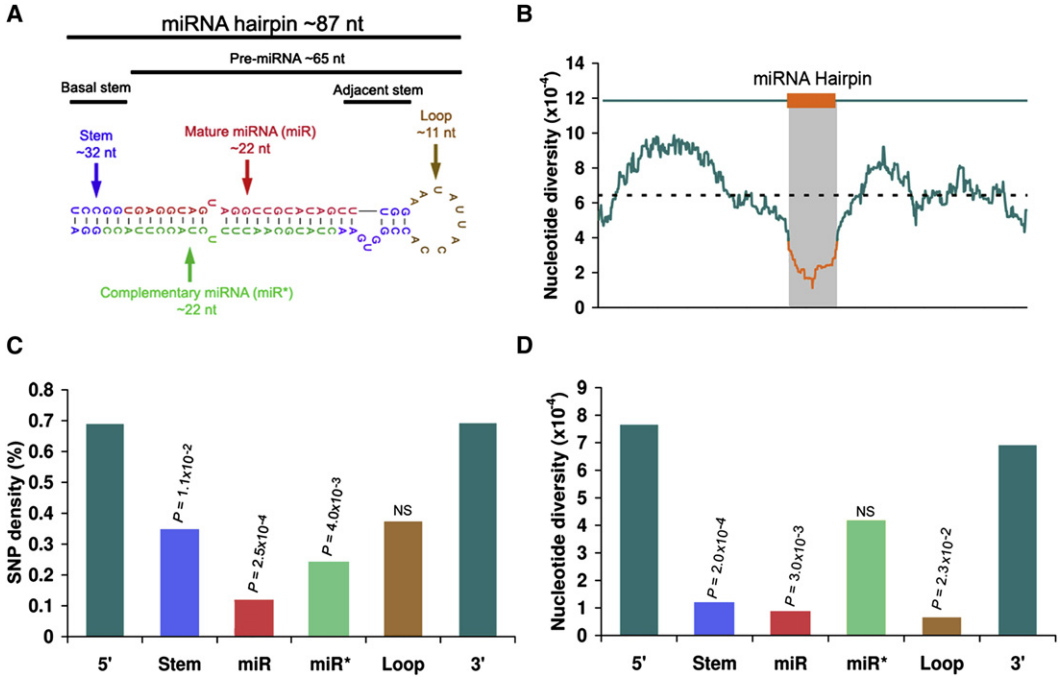

Figure 1. miRNA Hairpins Are Selectively Constrained in Human Populations

(A) The miRNA hairpin is composed of the precursor associated with its immediate flanking nucleotides that form a basal stem. The pre-miRNA is composed of an imperfect duplex between the mature miRNA (miR) and its complementary sequence (miR*), an adjacent stem, and a terminal loop. In our study, the stem refers to the basal stem in addition to the adjacent stem, excluding miR and miR*. The sizes considered correspond to the empirical mean sizes observed in our dataset for each of these different domains.

(B) Sliding windows of the nucleotide diversity π in miRNA-containing regions. All these regions were aligned on the first nucleotide of the mature miRNA. The window is 87 nt across (the mean size of the miRNA hairpins), with a step size of 1 nt. Each 87 nt window reflects the mean nucleotide diversity across all miRNA-containing regions. The orange line indicates significantly lower-than-expected levels of diversity over the sequenced regions, as determined by the resampling procedure (Material and Methods).

(C) Mean SNP density in the various miRNA domains.

(D) Mean nucleotide diversity π of the various miRNA domains. Both the mean SNP density and nucleotide diversity levels were calculated for the actual sizes of each of the different domains within the miRNA hairpins studied. p values were calculated with resampling analyses (Material and Methods). NS = non significant.

In order to test the impact of the simulated demographic model on the number of miRNA regions predicted to deviate from neutrality, we next considered a published demographic model that used a different set of noncoding regions in a different set of populations.[37] We chose this model because it is also based on full resequencing data from a large panel of noncoding regions (and thus avoids the potential problem of SNP ascertainment bias) in set of populations similar to ours (i.e., African, European, and Asian). We thus used the range of demographic parameters described in the Voight et al. study:[37] these are based on 50 unlinked noncoding genomic regions resequenced in 45 individuals from three human populations (15 Hausa, 15 central Italians, and 15 Han Chinese). We simulated non-African bottlenecks, conditionally on our European and East Asian sample sizes (23 and 20 individuals, respectively), by using their combination of parameters—i.e., a bottleneck starting 40,000 YBP in an ancestral population of 9,450 individuals with combinations of bottleneck duration and severity corresponding to the confidence region of parameter space with p values of 0.05 (Figures 2A and 2D in Voight et al.[37]). In addition, we also used their combination of parameters to simulate an African expansion, conditionally on our African sample size (48 individuals) — i.e., combinations of growth onset and growth rate of the exponential expansion corresponding to the confidence region of parameter space with *P*-values of 0.05 (Supporting Figure 3 in Voight et al.[37]). Finally, we re-estimated the p values of the neutrality tests for the miRNA regions corrected by this external demographic model[37] by counting the number of simulated values of the different summary statistics that did not fall into the range of observed values (Table S5).

Results

Levels of Naturally Occurring Sequence Variation in miRNAs Worldwide

We first investigated the naturally occurring sequence variation in human miRNAs by resequencing 92 genomic regions, covering 117 miRNA hairpins (Table S1), in 91 individuals from sub-Saharan Africa, Europe, and East Asia. We used the ~1.2 kb of sequence information surrounding each of the miRNA hairpins, which corresponds to a total of ~110 kb of diploid material per individual, to compute several summary statistics (Table S3). The global mean nucleotide diversity, π, was two times lower in miRNA-containing regions than in twenty

Table 1. SNPs Identified in miRNA Hairpins

miRNA						Frequency (%)[a]		
HUGO	Aliases	Location	Nt.[b]	dbSNP #	Alleles[c]	Africa	Europe	Asia
MIR105-2	miR-105-2	mirR	15	ss107938236	T/A	-	3.57	-
MIR220A	miR-220a	mirR	20	ss107938237	T/C	17.91	-	-
MIR379	miR-379	mirR	17	ss107938238	G/A	1.04	-	-
MIR146A	miR-146a	mirR*	40	rs2910164	G/C	36.46	23.91	55
MIR196A2	miR-196a-2	mirR*	54	rs11614913	C/T	16.67	39.13	47.5
MIR220A	miR-220a	mirR*	56	ss107938239	C/T	-	3.57	-
MIR339	miR-339	mirR*	50	ss107938240	A/G	-	2.17	-
MIR92-1	miR-92-1	mirR*	−27	ss107938241	C/T	1.04	-	-
MIR92-1	miR-92-1	mirR*	−26	rs9589207	G/A	8.33	-	-
MIR130B	miR-130b	Loop	−9	ss107938242	G/A	1.04	-	-
MIR34A	miR-34a	Loop	34	ss107938243	G/A	1.04	-	-
MIR93	miR-93	Loop	32	ss107938244	C/T	1.04	-	-
MIR96	miR-96	Loop	28	rs41274239	T/C	-	2.17	-
MIR222	miR-222	Loop	−9	ss107938245	G/A	2.99	-	-
MIR16-1	miR-16-1	Stem	42	ss107938246	T/C	4.17	-	-
MIR106B	miR-106b	Stem	35	ss107938247	G/T	2.08	-	-
MIR10A	miR-10a	Stem	−1	ss107938248	A/G	1.04	-	-
MIR124A2	miR-124a-2	Stem	43	ss107938249	A/G	-	-	2.5
MIR325	miR-325	Stem	−4	ss107938250	G/T	1.49	-	-
MIR339	miR-339	Stem	−8	ss107938251	G/A	-	-	5
MIR345	miR-345	Stem	−9	ss107938252	C/T	-	2.17	-
MIR183	miR-183	Stem	25	ss107938253	G/T	3.13	-	-
MIR215	miR-215	Stem	−14	ss107938254	T/C	-	2.17	-
MIR199B	miR-199b	Stem	−63	ss107938255	C/T	1.04	-	-
MIR216A	miR-216a	Stem	87	rs41291179	T/A	21.88	4.35	-
MIR492	miR-492	Stem	84	rs2289030	G/C	-	13.04	25
MIR194-2	miR-194-2	Stem	62	rs11231898	C/T	5.21	-	-

[a] The frequencies are relative to the derived state.
[b] Nucleotide position relative to the first site of the respective mature miRNA (miR).
[c] The first variant corresponds to the ancestral allele as compared to the chimpanzee sequence.

noncoding regions that were sequenced in the same individuals ($\pi = 0.68 \times 10^{-3}$ and $\pi = 1.16 \times 10^{-3}$, respectively, Mann-Whitney test, $p = 1.2 \times 10^{-3}$) (Tables S3 and S4). A similar pattern was observed when the continental population groups were considered separately (Mann-Whitney test, $p \leq 2.9 \times 10^{-3}$). In addition, the miRNA-containing regions generally exhibited lower levels of nucleotide diversity than those observed at different gene regions used for comparison (Figure S2). This resequencing effort allowed us to identify a total of 710 SNPs (data available upon request), 27 of which were located in the hairpin structure characterizing miRNAs: three SNPs in the mature miRNA sequence (the miR), six SNPs in the complementary miRNA sequence (the miR*), five SNPs in the loop, and 13 SNPs in the stem domain (Table 1). Although most SNPs within the hairpins were rare—20 out of the 27 SNPs exhibited a derived allele frequency lower than 5%—some of them were observed at high population frequencies. For example, within the mature sequence of miR-220a (*MIR220A*), we identified a SNP whose derived state is 18% frequency in Africans but absence elsewhere, and we identified two SNPs in the complementary sequences of miR-146a (*MIR146A*) and miR-196a-2 (*MIR196A2*), which presented derived allele frequencies ranging from 17%–55% in the different populations (Table 1).

Nucleotide Diversity Varies among the Different miRNA Domains

To test whether the generally reduced diversity observed in the miRNA-containing regions resulted from purifying selection (i.e., selective removal of deleterious variants) or simply reflected the presence of mutational cold spots in these regions, we performed a sliding-window analysis of the nucleotide diversity π over the different miRNA-containing regions. If reduced diversity were due to mutational cold spots alone, we would expect a random fluctuation of π over the sequenced regions. In contrast, purifying selection should lead to a reduced diversity specifically targeting the functional domains of the sequence, namely the miRNA harpins (Figure 1A). Strikingly, the 87 nt sliding-window analysis, corresponding to the mean length of the studied hairpins, showed that the reduced diversity was essentially due to a prominent and significant decay in the level of nucleotide diversity within the region corresponding to the miRNA hairpins (Figure 1B). This observation indicates that the miRNA hairpin structure has been subject to strong evolutionary constraint in human populations. The strong selective constraint of the hairpin structure observed in humans is further supported by the high levels of conservation of these miRNA hairpins across 28 vertebrate species

(Figure S3) and, more generally, by the strong conservation generally observed at miRNAs in primate cross-species comparisons.[23] Because the different domains composing the hairpin have been well characterized,[7,41] we next evaluated whether the level of sequence diversity fluctuated among them (i.e., miR, miR*, loop and stem). Indeed, some domains are expected to be more important for function than others. For example, the miR is the main actor in repressing activity of its mRNA targets, whereas the miR* has typically been assumed to be a mere carrier strand because of its short life compared to that of the mature strand during miRNA maturation.[11] Thus, the regional partitioning of function within the miRNA hairpins may result in distinct evolutionary dynamics among the different domains.

We thus calculated the SNP density and the nucleotide diversity π for each domain separately. The miR was the most highly constrained domain and displayed the lowest levels of SNP density ($p = 2.5 \times 10^{-4}$) (Figure 1C) and nucleotide diversity ($p = 3.0 \times 10^{-3}$) (Figure 1D) (see Material and Methods for details on the assessment of statistical significance). Interestingly, an absolute lack of diversity was observed in the first 14 nt of the miR ($p < 1 \times 10^{-4}$)—a segment that included the seed region (the first 7–8 nt) thought to be critical for mRNA target recognition.[42,43] Within the miR*, few mutations (i.e., derived alleles) appeared to reach appreciable frequencies at the population level (Table 1), making the mean level of nucleotide diversity π not significantly different from that of 5′ and 3′ flanking regions (Figure 1D). However, a general and significant lack of mutations was observed in this domain (Figure 1C). Interestingly, purifying selection has also affected the stem region, which shows both SNP density and nucleotide diversity to be significantly lower than expected (Figures 1C and 1D). For the loop domain, we did not observe a significant reduction in SNP density, but the mutations appeared to be maintained at very low frequencies ($p = 0.023$) (Figures 1C and 1D). Taken together, our results indicate that the miRNA hairpin structure is subject to strong selective constraint, especially in the miR domain, and are consistent with the miR's major role in mRNA target recognition. The other domains also seem to be constrained to varying extents, suggesting their potential function in the mechanism of mRNA translation repression.

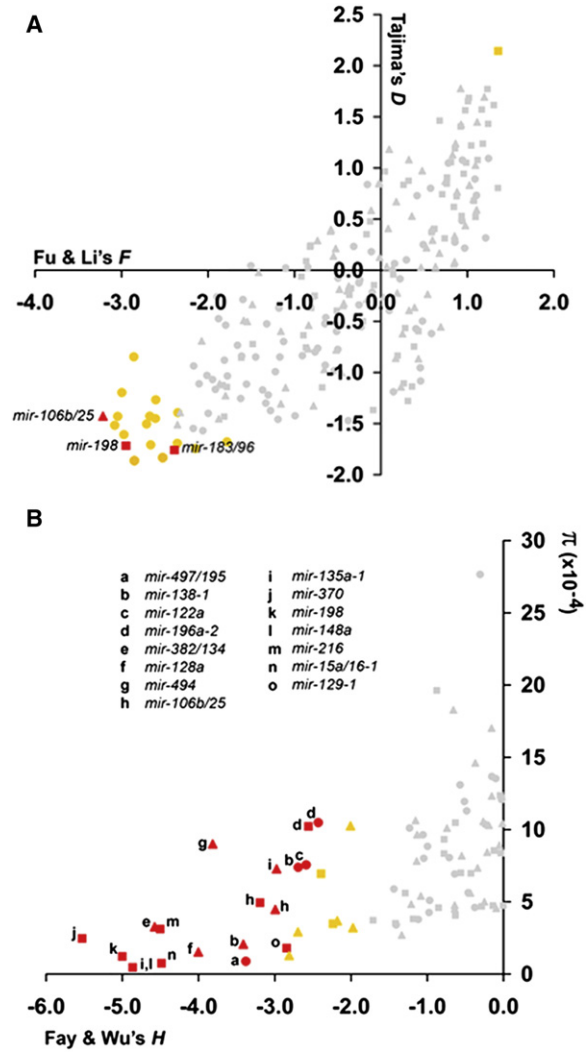

Figure 2. Sequence-Based Neutrality Tests in the Various Continental Populations Studied

(A) Tajima's *D* and Fu and Li's *F** in the African (circle), European (triangle), and East Asian (square) samples.

(B) Fay and Wu's *H* in the African (circle), European (triangle), and East-Asian (square) samples. miRNAs that are significant ($p < 0.025$ or $p > 0.975$ for two-sided tests, and $p < 0.05$ for the one-sided test) under a neutral model with constant population sizes are represented in yellow, and those that are significant after correction for the two demographic models are represented in red. Nonsignificant regions are shown in gray. Two-sided tests were used for Tajima's *D* and Fu and Li's *F**, and a one-sided test was used for Fay and Wu's *H*.

Testing for the Signatures of Other Selective Regimes: Sequence-Based Neutrality Tests

We then investigated whether other selective regimes, such as positive selection or balancing selection, had affected human miRNAs. We used several sequence-based neutrality tests (Tajima's *D*, Fu and Li's *F** and Fay and Wu's *H*) to evaluate whether the allele frequency spectrum of a given miRNA region deviated from expectations under neutrality.[44–47] The direction of these neutrality tests is potentially informative about the evolutionary and demographic forces that a population has experienced. For example, negative values of Tajima's *D* reflect an excess of rare polymorphisms in a population, which is consistent with positive or weak negative selection or an increase in population size. Positive values indicate instead an excess of intermediate-frequency alleles in a population and can result from balancing selection, population structure, or population bottlenecks.[44,46,47] We identified 47 miRNA-containing regions (i.e., 23 regions in Africans, 11 in Europeans, and 13 in East-Asians) whose variation was not compatible with neutrality (Figure 2 and Table S5). With

the exception of mir-101-1 (*MIR101-1*), which showed an excess of intermediate-frequency alleles in East Asians, all other regions were characterized by a significant excess of rare alleles (e.g., significantly negative Tajima's *D*) and/or an excess of high-frequency derived variants, as revealed by Fay and Wu's *H* test (Figure 2 and Table S5). Altogether, these patterns are compatible with the action of positive selection in a population-specific manner. However, because these commonly used neutrality tests for detecting selection make oversimplified assumptions (i.e., constant population size and no population structure) about past population history, the observed departures from neutrality may be explained by demographic events rather than natural selection.[46,47]

Integrating the Confounding Effects of Human Demography into Selective Inference

To experimentally account for demographic influences on the robustness-of-neutrality tests for detecting selection, we resequenced 20 independent noncoding regions dispersed throughout the genome, for a total of 27 kb per individual, in the same panel of individuals sequenced for the miRNA regions (Table S2 and Material and Methods). We used these empirical data to identify the demographic model best explaining the patterns of diversity (e.g., the number of polymorphic sites *S*, nucleotide diversity π, Tajima's *D*, etc.) observed at the noncoding regions (Tables S4 and S6). In perfect agreement with previous observations, the African data were most consistent with a population expansion, and the non-African data were most consistent with the occurrence of a bottleneck.[37,40,48,49] We subsequently simulated data over a broad range of parameters to identify the particular combination of demographic parameters that most closely matched the summary statistics of our empirical data for the 20 noncoding regions. To this end, we used an ABC approach,[38,39] which has been shown to allow the estimation of multiple demographic parameters (Fagundes et al.[40] and G.L. et al., unpublished data). The resulting range of parameters—the confidence intervals computed from the posterior distributions of the onset and intensity of the African expansion and the non-African bottleneck (see Material and Methods for details)—were consistent with previous studies.[37,40]

We re-estimated the significance of the neutrality tests (Tajima's *D*, Fu and Li's *F**, and Fay and Wu's *H*) for all miRNA regions in each population by using the previously estimated demographic parameters, which consider the occurrence of a bottleneck in non-Africans and an expansion among Africans. After we incorporated this demographic model into our neutral expectations, only 24 out of the initial 47 miRNA-containing regions retained significant values for neutrality tests (Table S5). We next aimed to validate the robustness of this demography correction that was based on the demographic parameters inferred from the resequencing data of the 20 noncoding regions. To this effect, we considered an external demographic model previously validated with an independent set of 50 unlinked noncoding regions sequenced in other populations[37] (see Material and Methods for further details). When simulating the demographic parameters provided by the Voight et al.'s model,[37] we found that 23 out of the initial 47 miRNA regions retained significant values for neutrality tests (Table S5). Taken together, our results clearly indicate that considering the oversimplified model of a panmictic constant-sized population results in the identification of a substantial number of false-positive signatures of selection. This observation emphasizes the need for integrating the confounding effects of the varying demographic histories of human populations when detecting the footprints of natural selection.

These two independent corrections for demography identified virtually the same set of miRNA regions as representing significant deviations from neutrality. These deviations were always in the direction of an excess of rare alleles, a pattern compatible with the action of weak negative selection or positive selection and/or with high-frequency derived alleles, a pattern unique to positive-selection events. Most of these signals were not shared among ethnic groups, suggesting population-specific selection events. Because local positive selection tends to increase population differentiation,[46,50,51] we next assessed the levels of population differentiation, by using F_{ST},[52] at all miRNA regions. We identified eight miRNA regions displaying unexpectedly high levels of differentiation between Africans and Europeans, seven between Africans and East Asians, and 11 between Europeans and East Asians (Figure 3), providing independent evidence that positive selection has targeted certain miRNA regions. Finally, we conservatively defined "robustly selected" miRNA regions as those that were significant after the two demographic corrections and that were also significant for at least two independent tests of selection (Table 2). With the exception of the miR-183/miR-96 region (*MIR183/MIR96*), whose patterns of variation are compatible with weak negative selection among East Asians, all miRNA regions were characterized by both an excess of high-frequency derived alleles and high levels of population differentiation. This distinctive signature supports the notion that these miRNA regions have been targeted by population-specific events of positive selection.

Signatures of Positive Selection in a Small-RNA-Rich-Island

Among the miRNA regions showing robust signatures of positive selection (Table 2), we noted enrichment for miRNAs located in the same genomic region. This region, herein named "small-RNA-rich island," spans ~200 kb on chromosome 14 and is not in LD with any known gene. The small-RNA-rich island harbors 48 miRNAs and 41 small nucleolar RNAs (snoRNAs) (Figure 4). snoRNAs can target RNAs for methylation or pseudouridylation, modify ribosomal and small nuclear RNAs, and control epigenetic imprinting.[53] We observed both extreme population

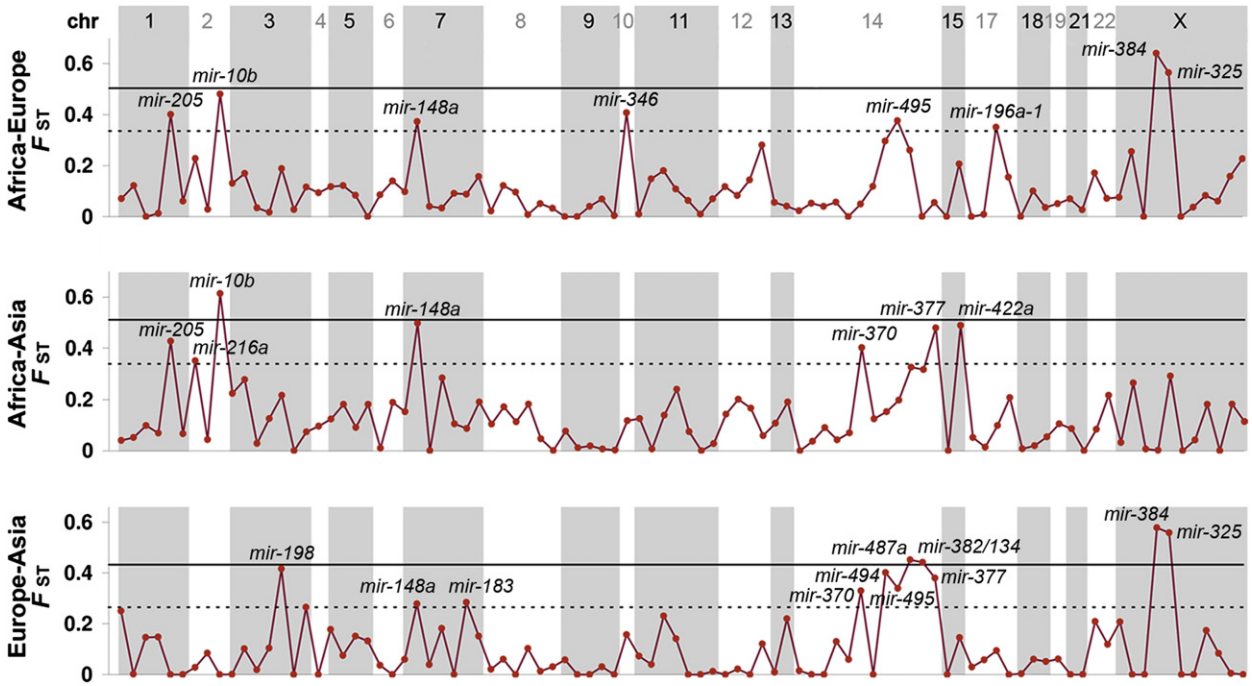

Figure 3. Multiple miRNA Regions Present High Levels of Population Differentiation
The dashed and solid lines correspond to the 95th and 99th percentiles obtained in simulations with our best-fitted demographic model. These thresholds are more stringent (i.e., higher) with respect to the 95th and 99th empirical values of F_{ST} resulting from the resequencing of the 20 noncoding regions (Table S8).

differentiation and an excess of high-frequenvy derived alleles for some miRNAs located in this region; such miRNAs included miR-494 (*MIR494*) and miR-382/miR-134 (*MIR382/MIR134*) (miRNA region A) in Europeans and miR-370 (*MIR370*) (miRNA region B) in East Asians (Table S5 and Figure 4). The absence of significant LD between regions A and B, regardless of the population considered (data not shown), indicates that independent events of positive selection might have occurred in Europeans and East Asians. We then investigated whether the small-RNA-rich island presented extended haplotype homozygosity, as expected under the effects of recent positive selection.[34,51] To this effect, we used HapMap phase II data[54] to obtain integrated Haplotype Score values (iHS)[34] for all SNPs genotyped in this region. This analysis localized more precisely the regions under positive selection; iHS values were highest in miRNA region A for Europeans and in the snoRNA region for East Asians (Figure 4). In East Asians, the signature of recent positive selection is among the strongest in the human genome; only 0.19% of the ~2 million polymorphic sites genotyped by HapMap in East Asians[54] presented iHS values higher than those observed for the snoRNA region.

Discussion

We have fully described the entire spectrum of sequence diversity characterizing 117 human miRNAs in different continental populations. This extensive resequencing effort allowed us to assess, without ascertainment bias, how this class of small RNAs has been targeted by natural selection during recent human evolution. Our results indicate that strong purifying selection affects the sequence corresponding to the mature miRNA. The strength of such selective constraint is particularly patent in the first 14 nt of the miR, where no mutation is tolerated. This observation, together with the strong selective constraint depicted in the miRNA target sites located in the 3′ UTR of human messenger RNAs,[55] emphasizes that a perfect Watson-Crick complementarity of the mature miRNA seed to its mRNA target is an essential parameter for repression activity. Interestingly, purifying selection has also affected the other domains of the miRNA hairpins to varying extents. Within the complementary miRNA sequence, despite the fact that some mutations can reach appreciable population frequencies, a significant lack of mutations was observed, indicating that most nucleotide positions are selectively constrained to ensure sufficient miRNA duplex pairing, an essential feature characterizing the miRNA hairpins. An additional factor most likely contributing to this observation is that, in some cases, the complementary miRNA functions as a genuine mature miRNA.[56] Therefore, some mutations occurring in the complementary strand may have measurable effects on gene regulation. The observed selective constraint in the adjacent stem region and in the loop supports the functional importance of these domains. Occurring mutations

Table 2. miRNA Regions Presenting Robust Signatures of Natural Selection

miRNA Region[a]		Tajima's D[b]			Fu and Li's F^*[b]			Fay and Wu's H[b]			Population Differentiation (F_{ST})[b]		
HUGO	Aliases	AF[c]	EU[c]	EAS[c]	AF	EU	EAS	AF	EU	EAS	EU/AF	EAS/AF	EU/EAS
MIR216A	**miR-216a**									−4.502**		0.352*	
MIR198	miR-198			−1.716*			−2.948**			−5.001**			0.416*
MIR148A	miR-148a									−4.862**	0.373*	0.497*	0.278*
***MIR106B, MIR93**, MIR25*	**miR-106b, miR-93**, miR-25					−3.214**			−2.993*	−3.191*			
MIR183, MIR96	**miR-183, miR-96**			−1.759*									0.284*
MIR370[d]	miR-370									−5.530**		0.402*	0.329*
MIR494[d]	miR-494								−3.814*				0.402*
MIR382,[d] *MIR134*[d]	miR-382, miR-134								−4.577*				0.442**

miRNAs in bold correspond to those in which SNPs have been identified in their hairpin structure (Table 1). The miRNA regions included in this table correspond to those that, after correction for both demographic models, kept rejecting neutrality for at least two independent tests of selection. These conservative criteria are intended to minimize the rate of false positive signatures of selection. The full list of miRNA regions rejecting neutrality for at least one neutrality test can be found in Table S5.

[a] miRNAs in the same line are organized in clusters.

[b] Significant results after correction for demography: *$p < 0.05$, ** $p < 0.01$.

[c] AF = Yoruba and Chagga from sub-Saharan Africa, EU = Danes from Europe, and EAS = Han Chinese from East Asia.

[d] The miRNAs contained in these regions are located in the so-called small-RNA-rich island on chromosome 14.

may be deleterious because of the changes in precursor secondary structure, hindering recognition by the Drosha-DGCR8 complex that is central to miRNA processing.[7,8]

The global signature of purifying selection depicted at the population level indicates that most miRNA mutations are selected against because of their deleterious consequences. However, such a strong selective constraint has not precluded an increase in the frequency of some miRNA mutations in the population by genetic drift or positive selection. Indeed, we have identified several miRNAs harboring mutations in their hairpin structure, and some of these were observed at high population frequencies (Table 1). One of these hairpin mutations, the G/C polymorphism within the miR* sequence of miR-146a (*MIR146A*), is known to exert a functional effect by reducing the amount of mature miR-146a and contributing to the genetic predisposition to papillary thyroid carcinoma in populations of European descent.[28] This mutation is present at ~24% of our European sample, in agreement with a previous study,[28] but it reaches even higher frequencies in the African and East Asian samples (36%–55%). The extent to which this mutation may contribute to the susceptibility to papillary thyroid carcinoma in African and East Asian populations remains to be investigated. Five additional miRNAs display mutations in their hairpins at particularly high frequencies; one of these is located in the mature sequence of miR-220a (*MIR220A*) and is found at ~18% frequency in the African sample. Interestingly, although this miRNA does not stand out among those presenting robust signatures of selection (Table 2) because of the highly conservative criteria we used to define them, Fu and Li's F^* test identified miR-220a (*MIR220A*) as being under balancing selection in Africa after correction by our demographic model (Table S5). Functional studies are now needed to determine whether the increased frequency of the miR-220a (*MIR220A*) mutation in Africa—and more generally of the remaining hairpin mutations identified in this study—reflects a selective advantage or, instead, neutral genetic drift.

Our study also provides evidence that local positive selection acts on a number of miRNA-containing regions. Sequence-based neutrality tests and population-differentiation levels allowed the identification of multiple miRNA regions showing robust signatures of population-specific positive selection after correction for the noise introduced by population demography. Nevertheless, several of these miRNA-containing regions lack mutations in the miRNA hairpin, and the few that were observed were mostly at low frequency. Thus, although the observed signs of positive selection could be due to LD with neighboring genes for some miRNAs (e.g., miR-198/*MIR198* is located in the 3′-UTR of *FSTL1*; Table S7), they could indeed reflect the fact that selection has targeted variants located in *cis*-regulatory regions of miRNAs. Interestingly, miR-148a (*MIR148A*), which presents one of the strongest selection signatures (Table 2), is not in LD with any known gene (Table S7) and has been shown to display high levels of interindividual variation in expression.[57] The expression of miR-148a (*MIR148A*) appears to be correlated with the translation efficiency of the pregnane X receptor (PXR [MIM 603065]), which regulates the expression of *CYP3A4* (MIM 124010).[57] CYP3A4 catalyzes the metabolism of more than 50% of commercially available drugs and is also subject to positive selection.[58] Natural selection may thus have favored the differential expression of miR-148a (*MIR148A*) as a means of controlling CYP3A4 synthesis via PXR. This example suggests that changes in the levels of miRNA expression may be of major functional and evolutionary relevance. In addition, our analyses revealed that a chromosome 14 genomic region—which is not in LD with any known gene and is highly enriched in miRNAs and snoRNAs—exhibits robust signatures of positive selection in Europeans and East Asians (Figure 4). Because no mutations were observed within the eight

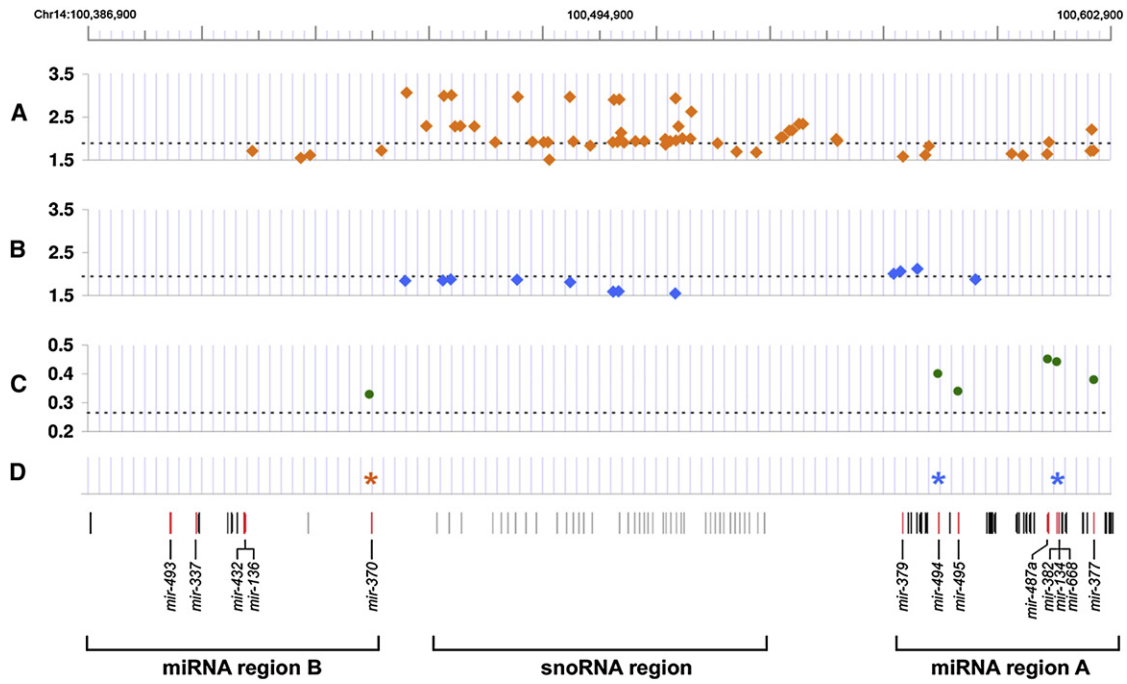

Figure 4. Recent Positive Selection Targeting a Small-RNA-Rich Island on Chromosome 14

Each line at the bottom of the figure represents a different small RNA; black and red lines denote miRNAs, and gray lines represent snoRNAs. Red lines indicate the miRNAs sequenced in this study.

(A and B) |iHs| values were calculated from HapMap II data for the SNPs genotyped in this region in (A) Asians and (B) Europeans. The dashed lines denote, for each population, the 95th percentile for the empirical value obtained from HapMap II data for chromosome 14. African |iHs| values are not reported because no significant deviation from neutral expectations was observed.

(C) F_{ST} values for pairwise comparisons of Europeans versus Asians. The dashed line denotes the 95th percentile obtained in simulations with our best-fitted demographic model (Table S8). Only significant values are reported.

(D) miRNAs presenting a positive-selection signature, as attested by a significant Fay and Wu's *H* value. The orange and blue stars indicate values that are significant after correction for demography in East Asians and Europeans, respectively.

miRNA hairpins sequenced in region A, the positive selection signature detected in Europeans probably corresponds to a selective event targeting either another miRNA or *cis*-regulatory elements of this region. In East Asians, our analyses identified a region enriched in snoRNAs under strong positive selection, suggesting that snoRNAs have also participated in the adaptive processes of current human populations. An extended resequencing effort in this region along with functional studies might help to bring to light the phenotypes driving selection on this small-RNA-rich island.

In conclusion, we have used full resequencing data of human miRNAs to show that this class of small RNAs has been strongly targeted by natural selection during recent human evolution. The signature of purifying selection globally shaping miRNA diversity worldwide indicates an essential and nonredundant role for these molecules in modulating gene-regulatory networks, and in human survival. This suggests that most mutations in miRNA hairpins are likely to be deleterious and therefore to have severe phenotypic consequences on human health. We also showed that some miRNA- and snoRNA-containing regions have been targeted by local positive selection. This observation should fuel future investigations exploring how genetic and functional variation, including differences in expression levels, of these small noncoding RNAs under selection affects the repression of their mRNA targets. This approach might identify miRNAs whose variation contributes to phenotypic diversity in humans and thus increase our understanding of the role of gene regulation in human adaptation, population differentiation, and complex disease etiology.

Supplemental Data

Supplemental Data include three figures and eight tables and are available with this article online at http://www.ajhg.org/.

Acknowledgments

We thank Howard Cann for critical reading of the manuscript. This work was supported by the Institut Pasteur, the Centre National de la Recherche Scientifique (CNRS), and the Agence Nationale de la Recherche (ANR) research grants (ANR-05-JCJC-0124-01 to L.Q.-M. and ANR-O6.BLAN-0061-02-AKROSS to C.A.). L.B.B. was supported by a "Fundação para a Ciência e a Tecnologia" fellowship (SFRH/BD/18580/2004) and E.P. by the Fondation pour la Recherche Médicale (FRM).

Received: November 10, 2008
Revised: December 21, 2008
Accepted: January 27, 2009
Published online: February 19, 2009

Web Resources

The URLs for data presented herein are as follows:

ALFRED, http://alfred.med.yale.edu
Arlequin v.3.11, http://cmpg.unibe.ch/software/arlequin3/
dbSNP NCBI, http://www.ncbi.nlm.nih.gov/SNP/
DnaSP v. 4.1, http://www.ub.es/dnasp/
GENALYS v. 3.3, http://software.cng.fr/
Haplotter, http://hg-wen.uchicago.edu/selection/haplotter.htm
miRBase Release 8.0, ftp://ftp.sanger.ac.uk/pub/mirbase/sequences/8.0/
Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim
PHASE v2.1.1, http://www.stat.washington.edu/stephens/software.html
SeattleSNPs project, http://pga.mbt.washington.edu/
SIMCOAL v. 2.0, http://cmpg.unibe.ch/software/simcoal/

Accession Numbers

The accession numbers of the newly identified SNPs in the miRNA hairpins presented in Table 1 are the following: ss107938236, ss107938237, and ss107938238 in the mature miRNA of miR-105-2 (*MIR105-2*), miR-220a (*MIR220A*), and miR-379 (*MIR379*), respectively; ss107938239, ss107938240, and ss107938241 in the complementary miRNA of miR-220a (*MIR220A*), miR-339 (*MIR339*), and miR-92-1 (*MIR92-1*), respectively; ss107938242, ss107938243, ss107938244, and ss107938245 in the loop of miR-130b (*MIR130B*), miR-34a (MIR34A), miR-93 (*MIR93*), and miR-222 (*MIR222*), respectively; and ss107938246, ss107938247, ss107938248, ss107938249, ss107938250, ss107938251, ss107938252, ss107938253, ss107938254, and ss107938255 in the stem of miR-16-1 (*MIR16-1*), miR-106b (*MIR106B*), miR-10a (*MIR10A*), miR-124a-2 (*MIR124A2*), miR-325 (*MIR325*), miR-339 (*MIR339*), miR-345 (*MIR345*), miR-183 (*MIR183*), miR-215 (*MIR215*), and miR-199b (*MIR199B*), respectively.

References

1. Ambros, V. (2004). The functions of animal microRNAs. Nature *431*, 350–355.
2. Bartel, D.P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. Cell *116*, 281–297.
3. Bartel, D.P., and Chen, C.Z. (2004). Micromanagers of gene expression: the potentially widespread influence of metazoan microRNAs. Nat. Rev. Genet. *5*, 396–400.
4. Lim, L.P., Glasner, M.E., Yekta, S., Burge, C.B., and Bartel, D.P. (2003). Vertebrate microRNA genes. Science *299*, 1540.
5. Lagos-Quintana, M., Rauhut, R., Lendeckel, W., and Tuschl, T. (2001). Identification of novel genes coding for small expressed RNAs. Science *294*, 853–858.
6. Du, T., and Zamore, P.D. (2005). microPrimer: the biogenesis and function of microRNA. Development *132*, 4645–4652.
7. Han, J., Lee, Y., Yeom, K.H., Nam, J.W., Heo, I., Rhee, J.K., Sohn, S.Y., Cho, Y., Zhang, B.T., and Kim, V.N. (2006). Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. Cell *125*, 887–901.
8. Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Radmark, O., Kim, S., et al. (2003). The nuclear RNase III Drosha initiates microRNA processing. Nature *425*, 415–419.
9. Hutvagner, G., McLachlan, J., Pasquinelli, A.E., Balint, E., Tuschl, T., and Zamore, P.D. (2001). A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. Science *293*, 834–838.
10. Ketting, R.F., Fischer, S.E., Bernstein, E., Sijen, T., Hannon, G.J., and Plasterk, R.H. (2001). Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans. Genes Dev. *15*, 2654–2659.
11. Lau, N.C., Lim, L.P., Weinstein, E.G., and Bartel, D.P. (2001). An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. Science *294*, 858–862.
12. Meister, G., Landthaler, M., Patkaniowska, A., Dorsett, Y., Teng, G., and Tuschl, T. (2004). Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. Mol. Cell *15*, 185–197.
13. Liu, J., Carmell, M.A., Rivas, F.V., Marsden, C.G., Thomson, J.M., Song, J.J., Hammond, S.M., Joshua-Tor, L., and Hannon, G.J. (2004). Argonaute2 is the catalytic engine of mammalian RNAi. Science *305*, 1437–1441.
14. Plasterk, R.H. (2006). Micro RNAs in animal development. Cell *124*, 877–881.
15. Valencia-Sanchez, M.A., Liu, J., Hannon, G.J., and Parker, R. (2006). Control of translation and mRNA degradation by miRNAs and siRNAs. Genes Dev. *20*, 515–524.
16. Croce, C.M., and Calin, G.A. (2005). miRNAs, cancer, and stem cell division. Cell *122*, 6–7.
17. Esquela-Kerscher, A., and Slack, F.J. (2006). Oncomirs - microRNAs with a role in cancer. Nat. Rev. Cancer *6*, 259–269.
18. Zhang, B., Pan, X., Cobb, G.P., and Anderson, T.A. (2007). microRNAs as oncogenes and tumor suppressors. Dev. Biol. *302*, 1–12.
19. Couzin, J. (2008). MicroRNAs make big impression in disease after disease. Science *319*, 1782–1784.
20. Lu, M., Zhang, Q., Deng, M., Miao, J., Guo, Y., Gao, W., and Cui, Q. (2008). An analysis of human microRNA and disease associations. PLoS ONE *3*, e3420.
21. Sethupathy, P., and Collins, F.S. (2008). MicroRNA target site polymorphisms and human disease. Trends Genet. *24*, 489–497.
22. Bentwich, I., Avniel, A., Karov, Y., Aharonov, R., Gilad, S., Barad, O., Barzilai, A., Einat, P., Einav, U., Meiri, E., et al. (2005). Identification of hundreds of conserved and nonconserved human microRNAs. Nat. Genet. *37*, 766–770.
23. Berezikov, E., Guryev, V., van de Belt, J., Wienholds, E., Plasterk, R.H., and Cuppen, E. (2005). Phylogenetic shadowing and computational identification of human microRNA genes. Cell *120*, 21–24.
24. Rigoutsos, I., Huynh, T., Miranda, K., Tsirigos, A., McHardy, A., and Platt, D. (2006). Short blocks from the noncoding parts of the human genome have instances within nearly all known genes and relate to biological processes. Proc. Natl. Acad. Sci. USA *103*, 6605–6610.
25. Griffiths-Jones, S., Saini, H.K., van Dongen, S., and Enright, A.J. (2008). miRBase: tools for microRNA genomics. Nucleic Acids Res. *36*, D154–D158.

26. Krek, A., Grun, D., Poy, M.N., Wolf, R., Rosenberg, L., Epstein, E.J., MacMenamin, P., da Piedade, I., Gunsalus, K.C., Stoffel, M., et al. (2005). Combinatorial microRNA target predictions. Nat. Genet. *37*, 495–500.
27. Lewis, B.P., Burge, C.B., and Bartel, D.P. (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell *120*, 15–20.
28. Jazdzewski, K., Murray, E.L., Franssila, K., Jarzab, B., Schoenberg, D.R., and de la Chapelle, A. (2008). Common SNP in pre-miR-146a decreases mature miR expression and predisposes to papillary thyroid carcinoma. Proc. Natl. Acad. Sci. USA *105*, 7269–7274.
29. Saunders, M.A., Liang, H., and Li, W.H. (2007). Human polymorphism at microRNAs and microRNA target sites. Proc. Natl. Acad. Sci. USA *104*, 3300–3305.
30. Takahashi, M., Matsuda, F., Margetic, N., and Lathrop, M. (2003). Automated identification of single nucleotide polymorphisms from sequencing data. J. Bioinform. Comput. Biol. *1*, 253–265.
31. Stephens, M., and Donnelly, P. (2003). A comparison of bayesian methods for haplotype reconstruction from population genotype data. Am. J. Hum. Genet. *73*, 1162–1169.
32. Rozas, J., Sanchez-DelBarrio, J.C., Messeguer, X., and Rozas, R. (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics *19*, 2496–2497.
33. Excoffier, L., Laval, G., and Schneider, S. (2005). Arlequin ver. 3.0: An integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online *1*, 47–50.
34. Voight, B.F., Kudaravalli, S., Wen, X., and Pritchard, J.K. (2006). A map of recent positive selection in the human genome. PLoS Biol. *4*, e72.
35. Wall, J.D. (1999). Recombination and the power of statistical tests of neutrality. Genet. Res. *74*, 65–79.
36. Laval, G., and Excoffier, L. (2004). SIMCOAL 2.0: A program to simulate genomic diversity over large recombining regions in a subdivided population with a complex history. Bioinformatics *20*, 2485–2487.
37. Voight, B.F., Adams, A.M., Frisse, L.A., Qian, Y., Hudson, R.R., and Di Rienzo, A. (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.
38. Beaumont, M.A., and Rannala, B. (2004). The Bayesian revolution in genetics. Nat. Rev. Genet. *5*, 251–261.
39. Beaumont, M.A., Zhang, W., and Balding, D.J. (2002). Approximate Bayesian computation in population genetics. Genetics *162*, 2025–2035.
40. Fagundes, N.J., Ray, N., Beaumont, M., Neuenschwander, S., Salzano, F.M., Bonatto, S.L., and Excoffier, L. (2007). Statistical evaluation of alternative models of human evolution. Proc. Natl. Acad. Sci. USA *104*, 17614–17619.
41. Seitz, H., and Zamore, P.D. (2006). Rethinking the microprocessor. Cell *125*, 827–829.
42. Brennecke, J., Stark, A., Russell, R.B., and Cohen, S.M. (2005). Principles of microRNA-target recognition. PLoS Biol. *3*, e85.
43. Lai, E.C. (2004). Predicting and validating microRNA targets. Genome Biol. *5*, 115.
44. Nielsen, R. (2005). Molecular signatures of natural selection. Annu. Rev. Genet. *39*, 197–218.
45. Sabeti, P.C., Schaffner, S.F., Fry, B., Lohmueller, J., Varilly, P., Shamovsky, O., Palma, A., Mikkelsen, T.S., Altshuler, D., and Lander, E.S. (2006). Positive natural selection in the human lineage. Science *312*, 1614–1620.
46. Nielsen, R., Hellmann, I., Hubisz, M., Bustamante, C., and Clark, A.G. (2007). Recent and ongoing selection in the human genome. Nat. Rev. Genet. *8*, 857–868.
47. Kreitman, M. (2000). Methods to detect selection in populations with applications to the human. Annu. Rev. Genomics Hum. Genet. *1*, 539–559.
48. Akey, J.M., Eberle, M.A., Rieder, M.J., Carlson, C.S., Shriver, M.D., Nickerson, D.A., and Kruglyak, L. (2004). Population history and natural selection shape patterns of genetic variation in 132 genes. PLoS Biol. *2*, e286.
49. Pluzhnikov, A., Di Rienzo, A., and Hudson, R.R. (2002). Inferences about human demography based on multilocus analyses of noncoding sequences. Genetics *161*, 1209–1218.
50. Barreiro, L.B., Laval, G., Quach, H., Patin, E., and Quintana-Murci, L. (2008). Natural selection has driven population differentiation in modern humans. Nat. Genet. *40*, 340–345.
51. Sabeti, P.C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., Xie, X., Byrne, E.H., McCarroll, S.A., Gaudet, R., et al. (2007). Genome-wide detection and characterization of positive selection in human populations. Nature *449*, 913–918.
52. Weir, B.S., and Cockerham, C.C. (1984). Estimating F-statistics for the analysis of population structure. Evolution *38*, 1358–1370.
53. Kiss, T. (2002). Small nucleolar RNAs: An abundant group of noncoding RNAs with diverse cellular functions. Cell *109*, 145–148.
54. Frazer, K.A., Ballinger, D.G., Cox, D.R., Hinds, D.A., Stuve, L.L., Gibbs, R.A., Belmont, J.W., Boudreau, A., Hardenbol, P., Leal, S.M., et al. (2007). A second generation human haplotype map of over 3.1 million SNPs. Nature *449*, 851–861.
55. Chen, K., and Rajewsky, N. (2006). Natural selection on human microRNA binding sites inferred from SNP data. Nat. Genet. *38*, 1452–1456.
56. Okamura, K., Phillips, M.D., Tyler, D.M., Duan, H., Chou, Y.T., and Lai, E.C. (2008). The regulatory activity of microRNA* species has substantial influence on microRNA and 3′ UTR evolution. Nat. Struct. Mol. Biol. *15*, 354–363.
57. Takagi, S., Nakajima, M., Mohri, T., and Yokoi, T. (2008). Post-transcriptional regulation of human pregnane X receptor by micro-RNA affects the expression of cytochrome P450 3A4. J. Biol. Chem. *283*, 9674–9680.
58. Thompson, E.E., Kuttab-Boulos, H., Witonsky, D., Yang, L., Roe, B.A., and Di Rienzo, A. (2004). CYP3A variation and the evolution of salt-sensitivity variants. Am. J. Hum. Genet. *75*, 1059–1069.