Maritime route of colonization of Europe

Peristera Paschou[a,1], Petros Drineas[b,1], Evangelia Yannaki[c], Anna Razou[d], Katerina Kanaki[d], Fotis Tsetsos[a], Shanmukha Sampath Padmanabhuni[a], Manolis Michalodimitrakis[d], Maria C. Renda[e], Sonja Pavlovic[f], Achilles Anagnostopoulos[c], John A. Stamatoyannopoulos[g], Kenneth K. Kidd[h], and George Stamatoyannopoulos[g,2]

[a]Department of Molecular Biology and Genetics, Democritus University of Thrace, 68100 Alexandroupolis, Greece; [b]Department of Computer Science, Rensselaer Polytechnic Institute, Troy, NY 12180; [c]Department of Hematology, George Papanicolaou Hospital, 57010 Thessaloniki, Greece; [d]Department of Forensic Medicine, University of Crete, Heraklion, 711 13 Crete, Greece; [e]Unità di Ricerca P. Cutino, Ospedali Riuniti Villa Sofia-Cervello, 90146 Palermo, Italy; [f]Institute of Molecular Genetics and Genetic Engineering, University of Belgrade, 11010 Belgrade, Serbia; [g]Departments of Medicine and Genome Sciences, University of Washington, Seattle, WA 98195; and [h]Department of Genetics, Yale University School of Medicine, New Haven, CT 06511

The Neolithic populations, which colonized Europe approximately 9,000 y ago, presumably migrated from Near East to Anatolia and from there to Central Europe through Thrace and the Balkans. An alternative route would have been island hopping across the Southern European coast. To test this hypothesis, we analyzed genome-wide DNA polymorphisms on populations bordering the Mediterranean coast and from Anatolia and mainland Europe. We observe a striking structure correlating genes with geography around the Mediterranean Sea with characteristic east to west clines of gene flow. Using population network analysis, we also find that the gene flow from Anatolia to Europe was through Dodecanese, Crete, and the Southern European coast, compatible with the hypothesis that a maritime coastal route was mainly used for the migration of Neolithic farmers to Europe.

Genotyping of extant and ancient populations has been used to address the question of the origins of the people of Europe. The genome of the present-day Europeans reflects merging of the Paleolithic settlers who colonized Europe 35,000–40,000 y before the present era (BPE) and the Neolithic people who started colonizing Europe approximately 9,000 y BPE. The Neolithic contribution to the gene pool of modern Europeans has been estimated with studies of extant European populations by using mitochondrial DNA, Y-chromosomal DNA, or nuclear DNA polymorphisms. Mitochondrial DNA studies estimate the Neolithic contribution to the maternal lineages of the modern Europeans to range between 10 and 20% (1). A contribution of approximately 22% was suggested by a study of Y-chromosome polymorphisms, which also found that the Neolithic contribution was more pronounced along the Mediterranean coast (2). Neolithic contributions of 50–70% were estimated with other methodologies (3–5), including highly polymorphic DNA markers (6). Clinal patterns of genetic diversity of autosomal (7–9) or Y-chromosomal (10) polymorphisms across Europe suggest that the Neolithic migrants originated from the Near East (7–9). It has been proposed that these Near Eastern migrants brought to Europe their new agricultural technologies (7–9, 11) and, perhaps, the Indo-European language (12). How did these Neolithic people reach Europe from the Near East?

The geographic focus of the transition from foraging to the Neolithic way of life was the Levantine corridor, which extended from the Fertile Crescent to the southeastern sections of the central Anatolian basin (13). The Neolithic farmers could have taken three migration routes to Europe. One was by land to North-Eastern Anatolia and from there, through Bosporus and the Dardanelles, to Thrace and the Balkans (14, 15). A second route was a maritime route from the Aegean Anatolian coast to the Mediterranean islands and the coast of Southern Europe (12, 14–18). The third was from the Levantine coast to the Aegean islands and Greece (19). Navigation across the Mediterranean was active during the Early Neolithic and Upper Paleolithic (16–18) as illustrated by the finding of obsidian from the island of Milos in Paleolithic sites of the Greek mainland (19, 20) and the early colonization of Sardinia, Corsica, and Cyprus (18, 21–23).

If a maritime route was used by the Neolithic farmers who settled Europe, their first stepping stones into Europe were the islands of Dodecanese and Crete. The Dodecanese is very close to the Aegean coast of Anatolia, whereas the west-most Dodecanesean islands are very close to Crete. Crete hosts one of the oldest Neolithic settlements of Europe in the site of Knossos, established ~8,500–9,000 y BPE (24, 25), and the inhabitants of the island established the first advanced European civilization starting approximately 5,000 BPE.

To obtain insights on the question of migrations to Europe, we analyzed genome-wide autosomal single nucleotide polymorphisms (SNPs) from a dataset of 32 populations. This dataset includes population samples from the islands of Crete and Dodecanese, one from Cappadocia in Central Anatolia, three subpopulations from different regions of mainland Greece, 14 other populations from Southern and Northern Europe, five populations from the Near East, and seven from North Africa. In addition to established methods for genetics analysis, we use a population genetics network approach that can define pathways of gene flow between populations. Our data are compatible with the hypothesis that a maritime route connecting Anatolia and Southern Europe through Dodecanese and Crete was the main route used by the Neolithic migrants to reach Europe.

Results

Genes Mirror Geography Across the Mediterranean Basin. We first used principal components analysis (PCA) to visualize the genotypic distances between studied populations (Fig. 1; also see *SI Appendix*,

---

Significance

The question of colonization of Europe by Neolithic people of the Near East and their contribution to the farming economy of Europe has been addressed with extensive archaeological studies and many genetic investigations of extant European and Near Eastern populations. Here, we use DNA polymorphisms of extant populations to investigate the patterns of gene flow from the Near East to Europe. Our data support the hypothesis that Near Eastern migrants reached Europe from Anatolia. A maritime route and island hopping was mainly used by these Near Eastern migrants to reach Southern Europe.

GENETICS

Figs. S1 and S2 and Tables S1 and S2). Populations on the southern and northern coasts of the Mediterranean, appear to be separated by the geographic barrier of the Mediterranean Sea. The role of the Mediterranean Sea as a barrier for gene flow among populations was also supported by our analysis using the BARRIER software (26), which implements Monmonier's maximum difference algorithm (*SI Appendix, SI Methods and Results* and Fig. S3). In accordance with this finding, notice, in Fig. 1*B*, that the PCA distribution of the populations closely resembles the geographic map of the countries circling the Mediterranean Sea. On this PCA "map" of populations, the east coast of the Mediterranean Sea is appropriately occupied by the Palestinians and the Lebanese Druze. Yemenites and Bedouins branch out from the Mediterranean populations and are closer to the populations of the Near East. Fig. 2 further illustrates the considerable resemblance of the PCA projection of the genotypes on the 2D space to the geographic map of the European Mediterranean coast. The east to west cline from Near East and Anatolia across the southern Mediterranean coast fits with the hypothesis of early

population movements from the Near East to Europe (7–9). The populations of the European Mediterranean Coast connect with the Near East although Anatolia (Cappadocia). In fact, the closest populations to Anatolia are those of Crete and Dodecanese rather than the populations of the Balkans or Northern Greece.

When considering Europe, Anatolia, and the Near East using PCA (Fig. 1*C*), a clear gradient is again observed, with populations from Northern and Central Europe connecting to Anatolia and the Near East via Southern Europe and through the bridge of the islands of Dodecanese and Crete. Three population tests (f3 statistics) as described in ref. 27 did not provide any evidence of bidirectional admixture or population splits along this line of stepping stones connecting Anatolia to Southeastern Europe (*SI Appendix, SI Methods and Results*). The correlation of geographic coordinates of the Mediterranean populations to the top two principal components (as shown in *SI Appendix,* Fig. S2D) is very high; the Pearson correlation coefficient is 0.66 between latitude (north to south axis) and the first principal component, and 0.81 between longitude (east to west axis) and the second

Fig. 1. Genes mirror geography around the Mediterranean coast. (*A*) Geographic distribution of the populations included in this study. (*B*) Projection on top two principal components of samples from 25 populations genotyped on ~75,000 genome-wide autosomal SNPs. A clear cline is observed with Anatolia (Cappadocia) connected to Southern Europe through the bridge of the islands of the Dodecanese and Crete. Bedouins and Yemenites drift toward Central-South Asia. No apparent gene flow between Northern Africa and the Southern coast of Europe is observed. (*C*) Projection on top two principal components of samples from 30 populations genotyped on the same set of SNPs presented in *B*. Northern European populations have now been added; Bedouins and Yemenites were removed. The cline now continues through Central and Northern Europe.
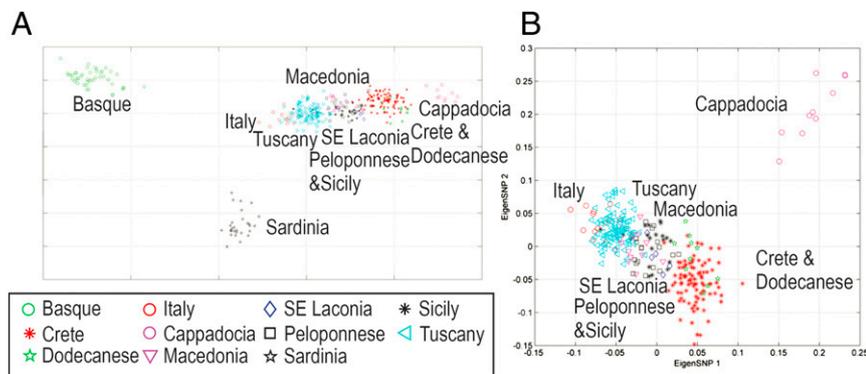
Fig. 2. Genetic structure of populations along the Southern European Coast in relation to Anatolia. (*A*) PCA plot of 10 populations from the Southern European coast and Cappadocia. The first principal component reveals the East to West cline in genetic variation along the Southern Coast of Europe and Mediterranean islands. Basques and Sardinians appear isolated relatively to the remaining studied populations. (*B*) Structure of the Southern European populations and Cappadocia excluding the more remote Basques and Sardinians.

principal component ($P < 10^{-5}$ as defined by Mantel test, see *SI Appendix*, Table S3 for geographic coordinates of the populations; this analysis does not even rotate or rescale the data, as was done in ref. 28). Thus, the PCA analysis supports a migration pathway from Anatolia to Europe through island hopping in the Aegean Sea.

To further analyze the relationships between the populations of the Mediterranean basin through an independent methodology, we used ADMIXTURE, an unsupervised ancestry-inference algorithm (Fig. 3; also see *SI Appendix*, Fig. S4 and S5). ADMIXTURE analysis on an LD-pruned dataset of 72,951 SNPs leads to essentially the same conclusions as PCA. The Near East and Anatolia are connected to Southern Europe through Crete and Dodecanese supporting the hypothesis of a pathway to Europe through the coastal Mediterranean route. ADMIXTURE analysis also illustrated the distinction between the populations of the North African coast and those of the European Mediterranean coast, and their connection through the populations of the Near East (Fig. 3). Thus, both the results of the PCA and the ADMIXTURE analysis are compatible with a maritime route of migrations from Anatolia/Near East to Southern Europe, in which Crete and the Dodecanese were used as the stepping stones by the migrating populations.

Network Analysis to Identify Pathways of Gene Flow Between Populations. To further test hypotheses on the routes of colonization of Southern Europe, we developed an approach for network-based population analysis. In doing so, we essentially attempted to reconstruct the pathways of gene flow between Near East, Anatolia, Mediterranean populations, and Northern/Central European populations as captured by PCA or ADMIXTURE. The goal was the creation of networks connecting the populations under study.

To form a network of related populations, we leveraged the fact that both PCA and ADMIXTURE essentially reduce the dimensionality of the original dataset by expressing each sample as either (*i*) a linear combination of the top few eigenvectors (in the case of PCA) or (*ii*) as percentages of ancestry from a small number of—typically unknown—ancestral populations (in the case of ADMIXTURE). To be more precise, in our study, each individual sample is originally described with respect to 75,194 SNPs; mathematically, this observation is equivalent to saying that the sample lies in a 75,194-dimensional subspace. After applying PCA or ADMIXTURE to the dataset, all samples are described with respect to K coefficients. K ranges between one and eight in all our results. Thus, the output of PCA or ADMIXTURE lies in a K-dimensional subspace, with K << 75,194. In the case of PCA, these coefficients correspond to the projections of each sample in the top K principal components (*SI Appendix*, Fig. S6), whereas in the case of ADMIXTURE, the coefficients correspond to the percentages for ancestry of each sample with respect to the K ancestral, but unknown, populations. It is worth indicating that PCA and ADMIXTURE are very different statistical techniques: PCA assumes the existence

of a small number of pairwise orthogonal components that explain the variance in the data, whereas ADMIXTURE is a model-based ancestry inference technique. (See *SI Appendix* for more details and a discussion of appropriate measures of distance for the low-dimensional PCA and ADMIXTURE subspaces.)

To form the networks, we first identified the top 10 nearest neighbors of each individual in the PCA or ADMIXTURE K-dimensional subspace, with the additional constraint that these nearest neighbors must not lie in the population of origin of the target individual. More specifically, we start by computing the ℓ_2 (Euclidean) distance (in the case of PCA) or the ℓ_1 (total variation) distance (in the case of ADMIXTURE) between the target individual and every other individual in our dataset. (See *SI Appendix* for a precise definition of the distance metric.) This procedure is repeated for all individuals in the dataset. Intuitively, our measure of distance dist(X,Y) between population X and population Y is the number of nearest neighbors that individuals in population X have in population Y. (A minor fine tuning can be used to mitigate the effect of different population sizes in our sample; see *SI Appendix* for details). The network formation algorithm can be described as follows: For each pair of populations X and Y, we compute both distances, dist(X,Y) and dist(Y,X). We create an edge between X and Y (thus claiming that the two populations are neighbors of each other) if min {dist(X,Y),dist(Y,X)} > 0, and we assign as a weight of the respective edge the value min{dist(X,Y),dist(Y,X)}. It is worth noting that our choice is quite conservative, because we assume that populations X and Y are related if, and only if, both X and Y have nearest neighbors in each other.

Population Network Analysis Supports a Maritime Path for the Colonization of Europe. Results of the network formation algorithm based on PCA are visualized with the Cytoscape software (Fig. 4*A*). In this figure, an edge between two populations shows that the two populations share genetically related individuals, with thicker edges indicating a larger number of genetically close individuals. The resulting networks (Fig. 4*A*; also see *SI Appendix*, Figs. S6 and S7) clearly indicate a path from the Near East populations (Palestinian, Druze) to Anatolia (Cappadocia), and from there to the islands of Dodecanese, and Crete. The connections between Crete and the rest of Greece (South East-Laconia → Peloponnese → Macedonia) as well as the populations of Sicily and Italy are evident. Analyses were performed by using the top three to seven PCs (*SI Appendix*, Figs. S6 and S7) and results were robust regardless of the number of principal components used. The geographic proximity and partial overlap in the PCA of Crete and Sicily is also compatible with gene flow from Crete to Italy and to Southern Europe through population movements along the Southern Mediterranean coast.

Results of our network analysis remain robust even when the independent methodology of ADMIXTURE is used to infer population distances. The resulting networks based on ADMIXTURE
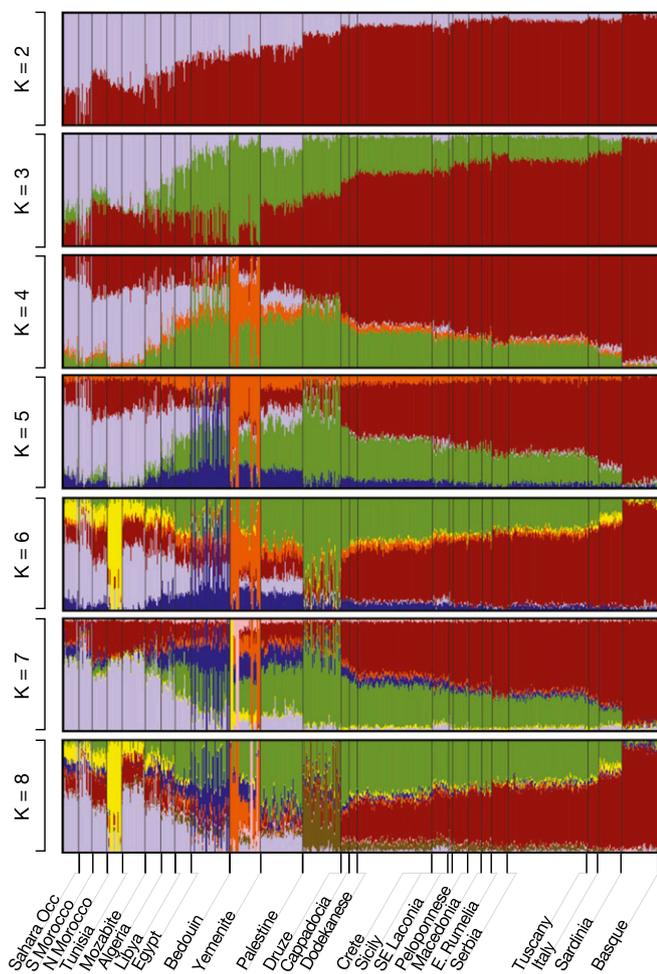
Fig. 3. Population structure around the Mediterranean basin. A model-based, unsupervised ancestry analysis approach (ADMIXTURE) was used to analyze populations on 72,951 (LD-pruned) genome-wide autosomal SNPs (*K* = 2–8). Two separate East to West clines are observed along the Northern African and Southern European Mediterranean coasts.
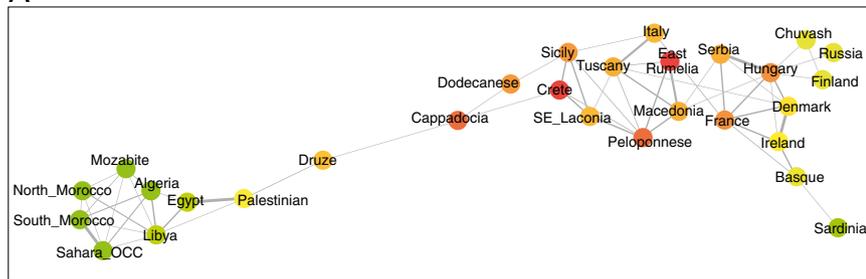
called closeness centrality with warmer colors representing more important nodes. Crete and Dodecanese are among the most central nodes in the network. More specifically, the closeness centrality is a classical network metric indicating the total distance from a node in a network to all other nodes. It is often regarded as a measure of how long it will take to spread information (in this case, genetic material) to other nodes in the network in a sequential manner. Crete is always the highest or the second highest ranking node in all networks that we formed with respect to this statistic, and Cappadokia is always among the top five nodes. Similarly, if we consider the so-called betweenness centrality, which quantifies the number of times a node in the network acts as a "bridge" along the shortest path between two other nodes, Crete and Cappadocia share the number one and two spots in almost all networks that we formed. East Rumelia, the Peloponnese, and the Dodecanese also rank among the top 12 nodes in most of the networks that we formed, indicating the pivotal role of the populations that were collected and included in our analysis in the migration from Near East and Anatolia to Europe.

Phylogenetic Analysis. We constructed a rooted neighbor-joining tree for our sample of 32 populations from North Africa, the Near East, Anatolia, and Southern and Northern Europe. The phylogenetic tree (*SI Appendix*, Fig. S12) was calculated from a matrix of Fst distances among populations, using the algorithms implemented in PHYLIP 3.695 and SmartPCA. Results are compatible with the clusters identified through PCA and ADMIXTURE analysis, and the identified lineages are largely concordant with the findings from our population network analysis. The populations from North Africa form a separate cluster branching out from South-West Asia. The closest European branches to the population of Cappadocia from Anatolia are those representing the islands of the Dodecanese and Crete. In turn, Crete connects to the most southern tip of mainland Greece (Southeast Laconia) as well as Sicily and the remaining populations from Northern and Central Europe that form a separate cluster. Thus, phylogenetic analysis also points to the central role of Crete and the Dodecanese in connecting Anatolia to Southern Europe. Similar results are obtained through analysis of our dataset with TreeMix (30) and Neighbor-Net (31). Both algorithms aim to construct a phylogenetic graph, improving the fit of a simple tree by allowing more than one path between populations. TreeMix and Neighbor-Net graphs underline gene flow from Anatolia to Southern Europe through the islands of the Dodecanese and Crete, without evidence for additional migrations from Anatolia through the Balkans (*SI Appendix*, Figs. S13 and S14 and Table S5).

Discussion

In historical times, there have been three major invasions of South Eastern Europe from the direction of the Near East but no evidence of major migratory events and gene flow. The Persians dominated South Western Asia in the fifth century BC: They established satrapies in Asia Minor and invaded Europe, but they were stopped by the Greeks (32). The Arabs attempted multiple invasions during the seventh and eighth centuries AD, but they were stopped by the Byzantines (33). An Arab tribe originating from Andalusia established in Crete a pirate state in the ninth century, but they were exterminated by the Byzantines 140 y later, and they left no traces of settlement in the island other than the name of their seat of power in the town of Chandax (33). The Turks invaded Asia Minor starting the 11th century and occupied the Balkans in the subsequent three centuries, but any Turks and converts to Islam left from Greek territories with the population exchanges that took place in the 20th century (34); the origin of the Turkish tribes was the central Asia. Seljuk Turks settled in Anatolia in the 12th century AD;
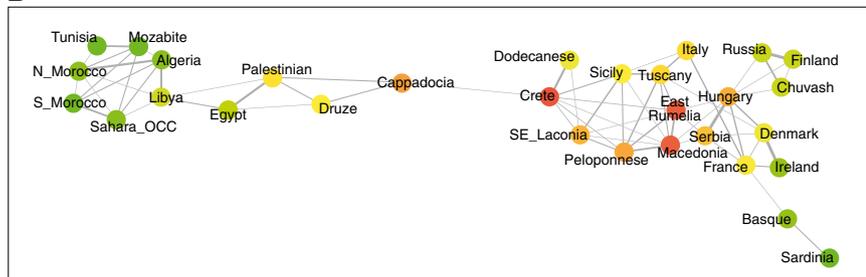
outputs are shown in Fig. 4*B* (also see *SI Appendix*, Fig. S8). Both the PCA and the ADMIXTURE-based networks tell a consistent story and the pathway of gene-flow from Anatolia to Southern Europe through the islands of Crete and Dodecanese is apparent. Results are robust, both with respect to the fact that two different approaches have been used (PCA and ADMIXTURE) and with respect to the dimensionality reduction parameter K, which varies between 3 and 8 for ADMIXTURE, and 3 and 7 for PCA. Network analysis using the Fst distance metric (*SI Appendix*, Fig. S9) does not provide very high resolution, but is similar in patterns to the networks obtained based on ADMIXTURE and PCA distances.

Simulations of a stepping stone model of migrations around the Mediterranean using IBDSim (29) (*SI Appendix, SI Methods and Results*) and investigation of simulated genotypes through PCA and our population network analysis (*SI Appendix*, Figs. S10 and S11 and Table S4), provide additional support to our hypothesis and leads to patterns that are strikingly similar to the actual observed patterns and relationships among populations, with Crete acting as a hub that connects Anatolia to Southeastern Europe.

Centrality Statistics Support the Role of Islands in Connecting Anatolia to Europe. As part of our network analysis using Cytoscape, we also computed centrality statistics of the nodes in our network. The color of the network nodes in Fig. 4 denotes the so-

Fig. 4. A coastal route of colonization of Europe. (*A*) A network analysis and visualization of the connections among 30 populations as revealed by the top five principal components. The network was formed by identifying nearest neighbors of each individual outside its populations of origin. Thicker edges represent stronger genetic relationships between pairs of populations, whereas warmer colors indicate high centrality of the respective nodes. The route connecting North Africa, Middle East, and Anatolia via the islands of the Dodecanese, and Crete to the rest of Europe, is apparent. (*B*) A network analysis and visualization of connections among 30 populations as revealed by ADMIXTURE with *K* = 5. Results are very similar to those in Fig. 4*A*, despite the fact that Admixture is a very different technique to extract ancestry information. Our networks are robust to the use of additional principal components or larger values of the ADMIXTURE parameter *K* for their formation (*SI Appendix*, Figs. S7 and S8).

however, the Anatolian Cappadocians we included in this study belong to the population that have kept the religion and the language of the pre-Seljuk Cappadocians and, therefore, most likely carry the genetic makeup of the ancient Anatolians. The only important gene flows from Near East to Europe must have occurred in prehistoric times and, as genetic evidence suggests, the most prominent migrations should have occurred during the Neolithic.

The idea that the Neolithic was introduced to Europe through coastal routes of colonization has been proposed by several archaeologists (12, 16, 17, 19, 22, 35). The earliest Neolithic sites with developed agricultural economies in Europe dated 8500–9000 BPE are found in Greece (19, 36, 37). The general features of material culture of the Greek Neolithic (14, 19, 36) and the genetic features of the preserved crops and associated weeds of the earliest Greek Neolithic sites point to Near Eastern origins (38). How these Near Eastern migrants reached Greece is a matter of speculation. One route of migration was by land from Central to Northeast Anatolia and from there to Southern Balkans through Bosporus, the Dardanelles, and Thrace (14, 15, 39). This migration route is less likely because archaeological evidence (19, 36, 40, 41) including ^{14}C dating (19, 40, 41) suggests that the Neolithic sites in Thrace and Macedonia are younger than those of mainland Greece, an unexpected finding if the Neolithic migrants who colonized Greece arrived there from the north. Other models suggest that waves of the Near-Eastern migrants reached Greece by sailing either from the Aegean Anatolian coast (12, 14, 16, 17, 22, 35) or from the Levantine coast (19, 36). Our data support the Anatolian rather than the Levantine route because they consistently show the Aegean islands to be connected to the Near East through Anatolia. Archaeological evidence from Greek and Near Eastern and Anatolian Neolithic sites suggests that multiple waves of Neolithic migrants reached Greece and Southern Europe. Most likely multiple routes were used in these migrations but, as our data show, the maritime route and island hopping was prominent. Our findings also suggest that to the west of Greece, the Neolithic reached Sicily and Italy by sea, as it has been suggested by archaeologists (12, 42).

Studies of extant European and Near Eastern populations using multiple autosomal genetic polymorphisms have established the presence of clinal distributions of allelic frequencies

(4–10, 43, 44). These clines in gene frequencies have been attributed to the geographically gradual merging of the gene pools of the Neolithic Near East migrants with the gene pools of the existing Paleolithic population of Europe. The correlation of clinal gene frequencies with the archaeological record of the spread of agriculture in Europe lead to the suggestion that it was the migration of Neolithic populations from the Near East that led to the spread of agriculture in Europe (7). The underlying hypothesis is that the development of agriculture triggered marked population growth and produced demographic pressures that resulted in dispersion of the Neolithic populations to new regions (7–9, 11). The rate of dispersion from the Near East to Western Europe has been estimated to approximately 0.6–1 km/y (44). A faster rate of dispersion is expected if maritime routes were used for the colonization of Southern Europe. Indeed, archaeological evidence suggests that farming spread faster in Southern Europe (12, 42, 45) and radiocarbon measurements in Neolithic sites are compatible with very rapid colonization of the west Mediterranean by Neolithic migrants (46, 47).

Although the Southeastern Mediterranean islands seem to have acted as a bridge from Anatolia to Southern Europe, the relatively small degree of gene flow between the African and the European coasts shows that the Mediterranean Sea also had a barrier function as also suggested with studies of mtDNA polymorphisms (48). Thus, the Mediterranean seems to have facilitated the migrations of Neolithic farmers along its Southern European coast but it mostly acted as an isolating factor between its European and African coasts.

Materials and Methods

GENETICS

released by Henn and coworkers (51), and (*iv*) data from the Kidd Laboratory at Yale University (allele frequencies for these data are available via the ALFRED database) (*SI Appendix*, Table S2).

PCA and ADMIXTURE. We used our own MatLab implementation of PCA (52, 53) (see *SI Appendix* for details). Before running ADMIXTURE, we pruned the SNPs to remove SNPs in high LD by using a windowed approach and a value of r^2 equal to 0.8.

Correlation Between PCA and Geographic Coordinates. We estimated the correlation between geographic coordinates (*SI Appendix*, Table S3) and the top two eigenvectors emerging from PCA. For each population in our sample, we approximated its location of origin either using information provided to us by the individuals that collected the respective sample, or by using a capital city that is relatively close to the population under study. The correlation between geographic coordinates and the eigenvectors was computed by converting both the geographic coordinates vector and the eigenvectors to *z* scores, and then computing the Pearson correlation coefficient. A Mantel test was run to estimate statistical significance.

Network Analysis. To better understand the connection between the populations included in our study, we performed a network analysis on the results of PCA and ADMIXTURE. To form the networks, we identified the top few nearest neighbors of each sample by representing each sample with respect to the top K coefficients returned by PCA or ADMIXTURE, and then computing the distance of each sample to all other samples, under the additional constraint that these neighbors should not belong to the same population of origin as the sample itself. Once a network whose nodes correspond to populations and whose edges correspond to connections between populations, as described above, is formed, we visualize it by using the Cytoscape software package (see *SI Appendix* for details).

1. Richards M, et al. (2000) Tracing European founder lineages in the Near Eastern mtDNA pool. *Am J Hum Genet* 67(5):1251–1276.
2. Semino O, et al. (2000) The genetic legacy of Paleolithic Homo sapiens sapiens in extant Europeans: A Y chromosome perspective. *Science* 290(5494):1155–1159.
3. Barbujani G, Bertorelle G (2001) Genetics and the population history of Europe. *Proc Natl Acad Sci USA* 98(1):22–25.
4. Cavalli-Sforza LL, Minch E (1997) Paleolithic and Neolithic lineages in the European mitochondrial gene pool. *Am J Hum Genet* 61(1):247–254.
5. Chikhi L, Nichols RA, Barbujani G, Beaumont MA (2002) Y genetic data support the Neolithic demic diffusion model. *Proc Natl Acad Sci USA* 99(17):11008–11013.
6. Belle EMS, Landry P-A, Barbujani G (2006) Origins and evolution of the Europeans' genome: Evidence from multiple microsatellite loci. *Proc Biol Sci* 273(1594): 1595–1602.
7. Menozzi P, Piazza A, Cavalli-Sforza L (1978) Synthetic maps of human gene frequencies in Europeans. *Science* 201(4358):786–792.
8. Ammerman AJ, Cavalli-Sforza LL (1984) *The Neolithic Transition and the Genetics of Populations in Europe* (Princeton Univ Press, Princeton, NJ), pp 85–108.
9. Sokal RR, Oden NL, Wilson C (1991) Genetic evidence for the spread of agriculture in Europe by demic diffusion. *Nature* 351(6322):143–145.
10. Rosser ZH, et al. (2000) Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. *Am J Hum Genet* 67(6):1526–1543.
11. Childe VG (1950) *The Dawn of European Civilization* (Routledge and Kegan Paul, London), pp 15–16.
12. Renfrew C (1987) *Archaeology and Language* (Jonathan Cape, London), pp 152–177.
13. Harris DR (1996) *The Origins and Spread of Agriculture and Pastoralism in Eurasia* (UCL Press, London), pp 552–573.
14. Hammond N (1976) *Migrations and Invasions in Greece and Adjacent Areas* (Noyes Press, Park Ridge, NJ), pp 84–99.
15. Özdoğan M (2011) Archaeological evidence on the westward expansion of farming communities from eastern Anatolia to the Aegean and the Balkans. *Curr Anthropol* 52(S4):S415–S430.
16. Broodbank C (2006) The origins and early development of Mediterranean maritime activity. *J Mediterr Archaeol* 19(2):199–230.
17. Davison K, Dolukhanov P, Sarson GR, Shukurov A (2006) The role of waterways in the spread of the Neolithic. *J Archaeol Sci* 33(5):641–652.
18. Cherry JF (1981) Pattern and process in the earliest colonization of the Mediterranean Islands. *Proc Prehist Soc* 47:41–68.
19. Perlès C (2001) *The Early Neolithic in Greece* (Cambridge Univ Press, Cambridge, UK), pp 58–62.
20. Runnels CN (1995) Review of Aegean prehistory IV: The stone age of Greece from the Palaeolithic to the advent of the Neolithic. *Am J Archaeol* 99(4):699–728.
21. Vigne J-D, et al. (2012) First wave of cultivators spread to Cyprus at least 10,600 y ago. *Proc Natl Acad Sci USA* 109(22):8445–8449.
22. Cherry JF (1990) The first colonization of the Mediterranean islands: A review of recent research. *J Mediterr Archaeol* 3(2):145–221.
23. Knapp AB (2013) *The Archaeology of Cyprus* (Cambridge World Archaeology, New York), pp 43–119.
24. Efstratiou N, Karestsou A, Banou ES, Margomenou D (2004) *Knossos: Palace, City, State*, eds Cadogan G, Hatzaki E, Vasilakis A (British School at Athens, London), pp 39–49.
25. Evans JD (1994) *Knossos: Labyrinth of History*, eds Evely D, Hughes-Brock H, Momigliano N (British School at Athens, London), pp 1–20.
26. Manni F, Guérard E, Heyer E (2004) Geographic patterns of (genetic, morphologic, linguistic) variation: How barriers can be detected by using Monmonier's algorithm. *Hum Biol* 76(2):173–190.

27. Patterson N, et al. (2012) Ancient admixture in human history. *Genetics* 192(3): 1065–1093.
28. Novembre J, et al. (2008) Genes mirror geography within Europe. *Nature* 456(7218): 98–101.
29. Leblois R, Estoup A, Rousset F (2009) IBDSim: A computer program to simulate genotypic data under isolation by distance. *Mol Ecol Resour* 9(1):107–109.
30. Pickrell JK, Pritchard JK (2012) Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet* 8(11):e1002967.
31. Bryant D, Moulton V (2004) Neighbor-net: An agglomerative method for the construction of phylogenetic networks. *Mol Biol Evol* 21(2):255–265.
32. Herodotus (1925) (Harvard Univ Press, Cambridge, MA).
33. Vasiliev AA (1952) *History of the Byzantine Empire* (Univ Wisconsin Press, Madison, WI), Vol 1, pp 199–218, 303–315.
34. (1925) Greece and Turkey - Convention concerning the exchange of Greek and Turkish populations and protocol. *League of Nations Treaty Series* 32:76–87.
35. Broodbank C, Strasser TF (1991) Migrant farmers and the Neolithic colonization of Crete. *Antiquity* 65(247):233–245.
36. Demoule J-P, Perlès C (1993) The Greek Neolithic: A new review. *J World Prehist* 7(4): 355–416.
37. Van Andel TH, Runnels CN (1995) The earliest farmers in Europe. *Antiquity* 69(264): 481–500.
38. Colledge S, Conolly J, Shennan S (2004) Archaeobotanical evidence for the spread of farming in the eastern Mediterranean. *Curr Anthropol* 45(Supp):S35–S58.
39. Özdoğan M (1997) The beginning of Neolithic economies in southeastern Europe: An Anatolian perspective. *J Eur Archaeol* 5(2):1–33.
40. Cauwe N, Dolukhanov P, Kozlowski J, Van Berg PL (2007) *Le Néolithique en Europe* (Armand Collin, Paris), pp 77–97.
41. Andreou S, Fotiadis M, Kotsakis K (1996) Review of Aegean Prehistory V: The Neolithic and Bronze Age of Northern Greece. *Am J Archaeol* 100(3):537–597.
42. Phillips P (1975) *Early Farmers of West Mediterranean Europe* (Hutchinson Univ Library, London), pp 44–74.
43. Sokal R, Menozzi P (1982) Spatial autocorrelations of HLA frequencies in Europe support demic diffusion of early farmers. *Am Nat* 119(1):1–17.
44. Pinhasi R, Fort J, Ammerman AJ (2005) Tracing the origin and spread of agriculture in Europe. *PLoS Biol* 3(12):e410.
45. Whittle A (1994) The first farmers. *The Oxford Illustrated Prehistory of Europe*, ed Cunliffe B (Oxford Univ Press, Oxford), pp 136–166.
46. Zilhão J (2001) Radiocarbon evidence for maritime pioneer colonization at the origins of farming in west Mediterranean Europe. *Proc Natl Acad Sci USA* 98(24): 14180–14185.
47. Zilhao Z (2000) From the Mesolithic to the Neolithic in the Iberian Peninsula. *Europe's First Farmers*, ed Price TD (Cambridge Univ Press, Cambridge, UK), pp 144–182.
48. Simoni L, Gueresi P, Pettener D, Barbujani G (1999) Patterns of gene flow inferred from genetic distances in the Mediterranean region. *Hum Biol* 71(3):399–415.
49. Li JZ, et al. (2008) Worldwide human relationships inferred from genome-wide patterns of variation. *Science* 319(5866):1100–1104.
50. The International HapMap 3 Consortium (2010) Integrating common and rare genetic variation in diverse human populations. *Nature* 467(7311):52–58.
51. Henn BM, et al. (2012) Genomic Ancestry of North Africans Supports Back-to-Africa Migrations. *PLoS Genetics* 8(1):e1002397.
52. Paschou P, et al. (2007) PCA-correlated SNPs for structure identification in worldwide human populations. *PLoS Genet* 3(9):1672–1686.
53. Paschou P, et al. (2007) Intra- and interpopulation genotype reconstruction from tagging SNPs. *Genome Res* 17(1):96–107.