

Geographic Population Structure (GPS) of worldwide human populations infers biogeographical origin down to home village

Eran Elhaik and Tatiana Tatarinova

The search for a biogeographical method that utilizes biological information to predict one's place of origin has occupied scientists for millennia. Modern biogeographical algorithms achieve an accuracy of 700 km in Europe but are highly inaccurate elsewhere, particularly in Southeast Asia and Oceania. Here, we present the admixture-based Geographic Population Structure (GPS) method that accurately infers the biogeography of worldwide individuals down to their village of origin. GPS's accuracy is demonstrated on three datasets: worldwide populations, Southeast Asians and Oceanians, and Sardinians (Italy) using 40,000-130,000 GenoChip markers. GPS correctly placed 80% of worldwide individuals within their country of origin with an accuracy of 87% for Asians and Oceanians. Applied to over 200 Sardinian villagers of both sexes, GPS placed a quarter of them within their villages and most of the remaining within 50 km of their villages, allowing us to identify the demographic processes that shaped the Sardinian society.

Finally, we demonstrate additional applications of GPS in tracing the biogeographical origin of the Druze population and uncovering the European origins of North Americans. The accuracy and power of GPS underscore the promise of admixture-based methods to biogeography and has important ramifications for genetic ancestry testing, forensic and medical sciences, and genetic privacy.