Supporting Information

Tatem and Smith 10.1073/pnas.1002971107

SI Text

Migration Network Communities Within *P. falciparum* Malaria Endemic Regions. Mapping "natural" migration boundaries that display reduced movement of people and infections between regions has practical utility, not just in the planning of malaria elimination. These enable interventions to be applied and coordinated in regional blocks afforded protection from recolonization by natural firebreak regions of reduced migration. Here, a recently constructed bilateral international migration database was analyzed with network analysis tools to infer, within the *P. falciparum* endemic regions of the world, communities of countries linked by strong population exchange, and those regions where movement levels are lower.

Identifying Migration Network Communities. Network analysis approaches were applied to the rescaled migrant database to identify well-defined communities of countries that are connected by high levels of population exchange. These communities were defined using random walk-based methodologies extended for weighted (1) and directed (2) networks, and the strength of community linkage evaluated using measures of network modularity (3). In simple terms, at each stage of the analysis, the two countries (or communities of countries) most strongly connected by population exchange were merged to form a new community. The strength of the new set of communities and individual countries in terms of population exchanges was then measured and recorded. These steps were repeated until all countries formed a single community. These steps are described in detail below.

The "walktrap" community structure algorithm was used to identify communities of countries that are most strongly linked by migration. The algorithm is based on the fact that random walks on a network tend to get trapped in strongly weighted parts corresponding to communities, and is described in detail in Pons and Latapy (1). A range of alternative algorithms exist for the detection of strongly connected communities within networks (3–5), and each of these were tested alongside the walktrap algorithm, but little change in results were found.

Network weights were calculated using methods outlined in Kim et al. (2) that account for directionality within a network. Initially, the set of n countries formed a weighted network of n communities. Two communities, C_1 and C_2, were then chosen according to:

$$\sigma_s = \frac{1}{n} \sum_{c \in p_s} \sum_{i \in c} d_{ic}^2.$$

At each step s, the two communities that minimize the mean σ_s of the squared distance (d) between each vertex i and its community C were merged. This created a new community C_Z and a new partition of the migrant database, P_{S+1}. Distances between communities were then updated and the procedure was repeated until $n-1$ steps and partition P_n was obtained. Each step defined a partition P_s of the graph into communities of countries, which provided a hierarchical structure of communities.

Quantifying and exploring which communities were the strongest in terms of migrant exchange was undertaken by calculating, for each partition P_s from the walktrap algorithm, a network modularity score, *Q(P)*. Modularity is a measure of the quality or strength of a given division of a network and is measured by

$$Q(p) = \sum_{c \in p} e_c - a_c^2.$$

The best partition is considered to be the one that maximizes Q, where e_c is the fraction of edges inside community C, and a_c is the fraction of edges bound to community C. Modularity is described in detail in Newman and Girvan (3). Here, the migrant community divisions for each world region that produced both the largest modularity scores and the largest increases in earlier steps between individual agglomeration steps were identified and mapped to visualize the strongest linkages within communities. All approaches were implemented using the igraph package (http://igraph.sourceforge.net/) within the R statistical computing environment (6).

Results Figs. S1–S3 show results of regional community structure analyses based on migration data, highlighting those countries that form communities linked by high levels of population exchange. In each case, a graph of the modularity scores for each step of agglomeration of countries into new communities is presented, along with maps showing the communities identified at significant points in the agglomeration process.

Fig. S1 shows that intercountry migration is stronger in some areas of the Africa and Arabian Peninsula region than others. Throughout the agglomeration steps shown, communities of countries in West Africa (in red) and Southern Africa (in green) are consistently found, highlighting the greater levels of migrant exchange and circulation between these countries than to the rest of the region. In contrast, other countries remain isolated, even at some of the highest levels of agglomeration. Madagascar, Ethiopia, and Saudi Arabia exhibit no distinct membership to a migration network community within the Africa and Arabian Peninsula area until the final agglomeration steps, indicating overall low levels of in- and out-migration to the remainder of the region.

The Americas display a range of community structures (Fig. S2), with the highest modularity score defining four communities: (*i*) the Central American *P. falciparum* malaria endemic countries, (*ii*) continental South America, (*iii*) Hispaniola, and (*iv*) Panama. The low modularity scores for earlier iterations shows that smaller communities of countries linked by strong migrant exchange are not such a feature of the Americas as was seen for the Africa and Arabian Peninsula region.

Fig. S3 shows that the community configuration for Central and Southeast (CSE) Asia that results in the largest modularity score is a division into three distinct and spatially contiguous regions representing West, Central, and East Asia. It is the earlier stages of the agglomeration process that reveals interesting within-region structures in terms of countries connected by high levels of migrant exchange. For instance, the first set of communities shown (Fig. S3, *Top Left*) highlights the relative isolation of Papua New Guinea, the Solomon Islands, and Vanuatu, with each displaying no community membership. In central East Asia it can be seen that Vietnam, Cambodia, Laos, and Thailand are strongly connected to each other by population exchange. In contrast, Myanmar is shown to be relatively isolated in terms of population migration to and from surrounding countries.

Migration and *P. falciparum* Malaria Sources and Sinks: Americas and Central and Southeast Asia. Figs. S4 and S5 show which nations in the Americas and Central and Southeast (CSE) Asia are likely net exporters of *P. falciparum* parasites and which are net importers. In combination with Figs. 2 and 3 of the main text, these maps describe general patterns of infection movements globally. Movements are categorized by their origin *Pf*PR$_{2-10}$ endemicity class, as defined by Hay et al. (7). Here we focus on examining the control-relevant situations of movements from high-transmission ($Pf\mathrm{PR}_{2\text{–}10} > 40\%$) to low-transmission regions ($Pf\mathrm{PR}_{2\text{–}10} < 40\%$).

The Americas are characterized by generally low *P. falciparum* transmission (7), thus international population movements bringing imported *Pf* infections represent less of a significant issue control-wise than for the Africa and Arabian Peninsula region. Nevertheless, Fig. S4 highlights where significant flows likely result in relatively large numbers of imported *P. falciparum* infections, requiring trans-border efforts in any control or elimination program. Two specific examples stand out: (*i*) the large numbers moving from Haiti with universal 0–5% $Pf\mathrm{PR}_{2\text{–}10}$ transmission, to the Dominican Republic, which is largely malaria-free or of unstable transmission and (*ii*) the high migration from Colombia to Venezuela, moving from a principally 0–5% $Pf\mathrm{PR}_{2\text{–}10}$ transmission country to a principally unstable transmission and malaria-free country.

CSE Asia (Fig. S5) is generally characterized by low *P. falciparum* transmission, but with pockets of medium-high transmission (7) and significant population movements in certain areas. In terms of movements from high- to low-transmission regions, Southeast Asia is shown here as being an area of substantial interest. Relatively high levels of outmigration from Cambodia and Laos, where the highest transmission classes predominate, to Thailand and Vietnam, where unstable transmission dominates, are evident. This is reflected in malaria incidence data from Vietnam and Thailand, where malaria has to a large extent become a border problem (8). This is also seen, although to a lesser extent, in Malaysia and Indonesia, with movement from surrounding countries.

1. Pons P, Latapy M (2005) Computing communities in large networks using random walks. *Lect Notes Comput Sci* 3733:284–293.
2. Kim Y, Son S-W, Jeong H (2010) Link rank: Finding communities in directed networks. *Phys Rev E Stat Nonlin Soft Matter Phys* 81:016103.
3. Newman MEJ, Girvan M (2004) Finding and evaluating community structure in networks. *Phys Rev E Stat Nonlin Soft Matter Phys* 69:026113.
4. Clauset A, Newman MEJ, Moore C (2004) Finding community structure in very large networks. *Phys Rev E Stat Nonlin Soft Matter Phys* 70:066111.
5. Reichardt J, Bornholdt S (2006) Statistical mechanics of community detection. *Phys Rev E Stat Nonlin Soft Matter Phys* 74:016110.
6. R Development Core Team (2009) R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing, Vienna). Available at http://www.R-project.org/. Accessed May, 2010.
7. Hay SI, et al. (2009) A world malaria map: *Plasmodium falciparum* endemicity in 2007. *PLoS Med* 6:e1000048.
8. Schapira A (2007) *Prospects for Eradication and Elimination of Malaria: Annexe 2. Opportunities, Obstacles and Risks for Elimination of Plasmodium falciparum Malaria in Different Countries and Regions of the World with Currently Existing Tools* (Department for International Development, London).

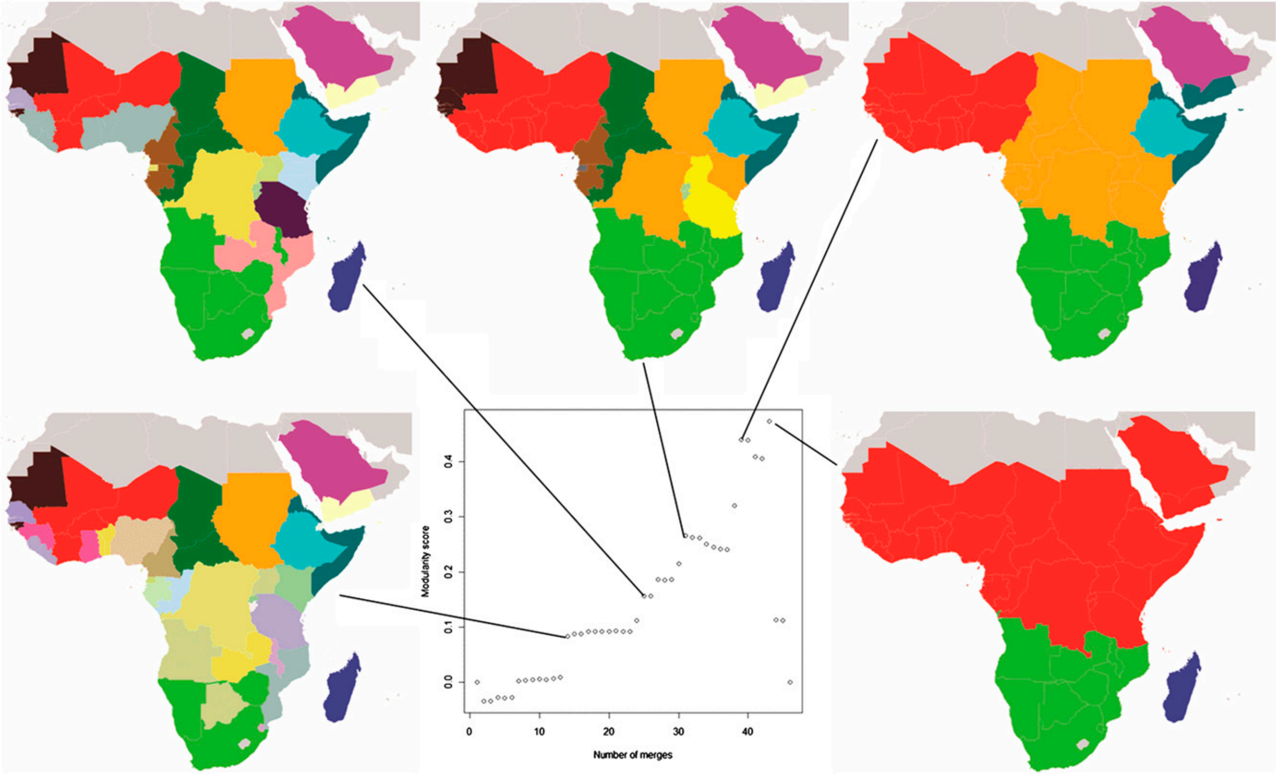

Fig. S1. Migration communities for the Africa and Arabian Peninsula region. Each map represents a different stage of country-merging into communities connected by high levels of migration, with community membership shown by color. For instance, the bottom-left map shows that Mali, Niger, Burkina Faso, and Côte d'Ivoire form a community (colored red) that is strongly linked by population exchange. The plot in the center shows the overall strength (measured by modularity score) of clustering at different stages of merging countries together into communities. The stage that each map shown represents is identified by the connecting lines.

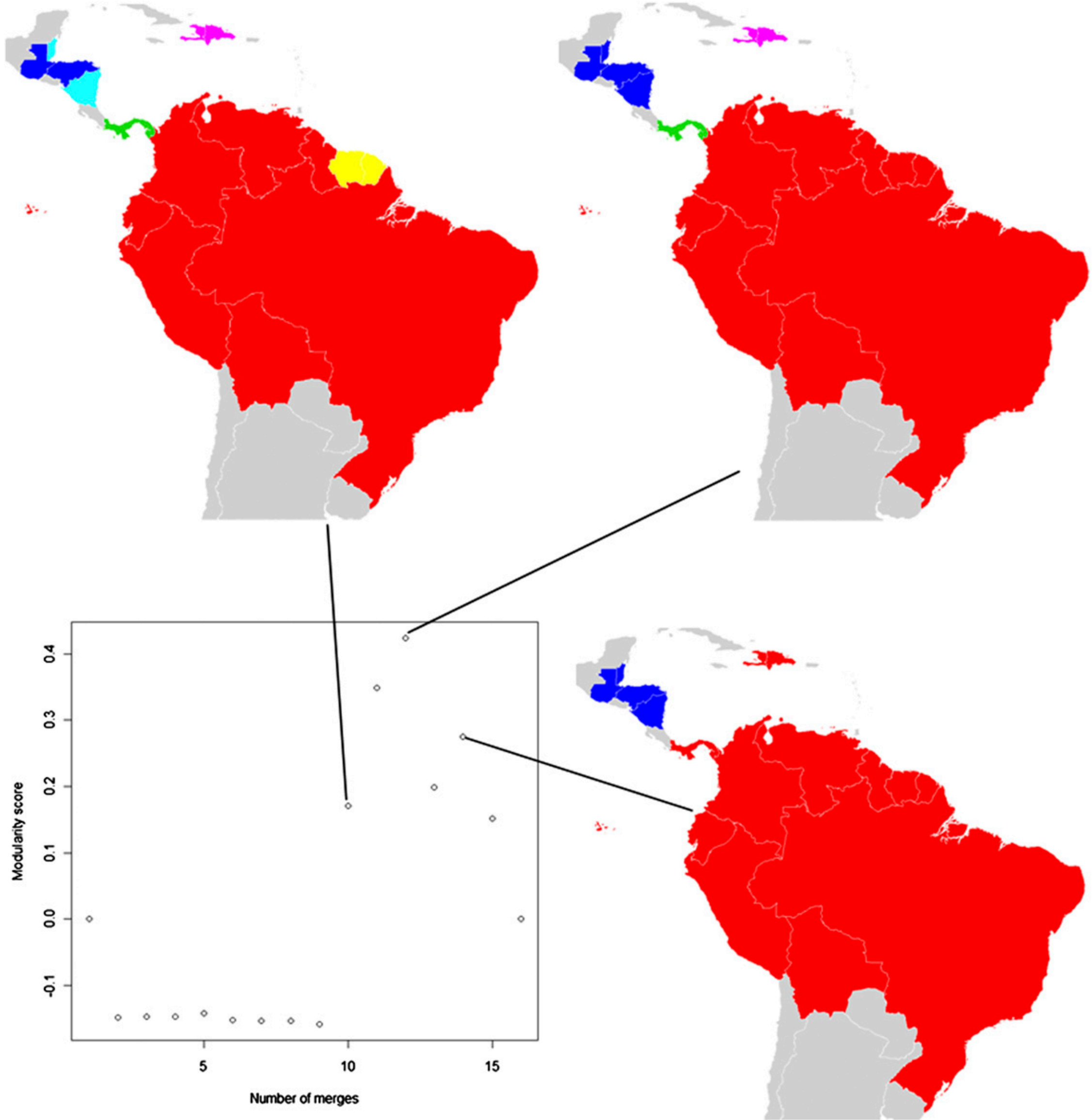

Fig. S2. Migration communities for the Americas region. Each map represents a different stage of country-merging into communities connected by high levels of migration, with community membership shown by color. For instance, the top-left map shows that the Dominican Republic and Haiti form a community (colored pink) that is strongly linked by population exchange. The plot in the center shows the overall strength (measured by modularity score) of clustering at different stages of merging countries together into communities. The stage that each map shown represents is identified by the connecting lines.

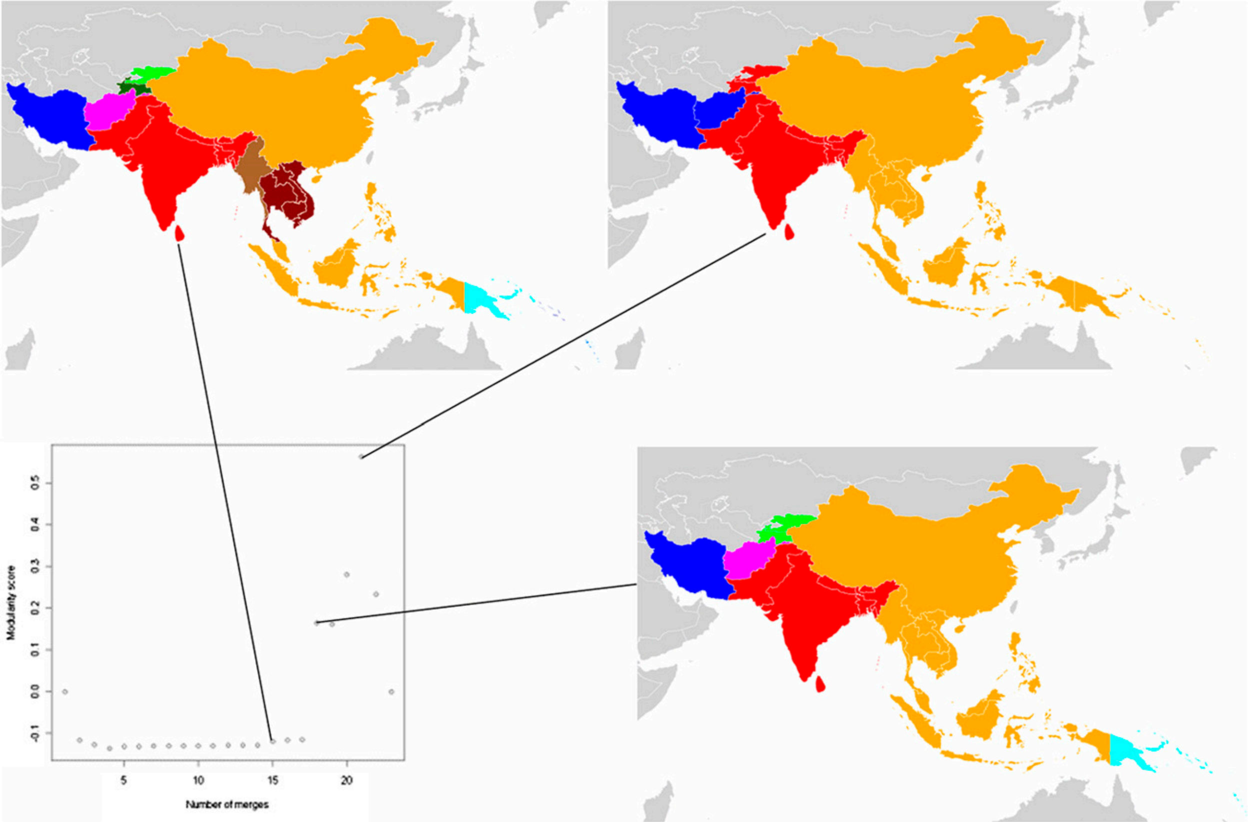

Fig. S3. Migration communities for the Central and Southeast Asia region. Each map represents a different stage of country-merging into communities connected by high levels of migration, with community membership shown by color. For instance, the top-left map shows that Thailand, Vietnam, Cambodia, and Laos form a community (colored dark brown) that is strongly linked by population exchange. The plot in the center shows the overall strength (measured by modularity score) of clustering at different stages of merging countries together into communities. The stage that each map shown represents is identified by the connecting lines.

Fig. S4. Americas: (*A*) outgoing migrants and (*B*) incoming migrants by endemicity class. Endemicity classes are colored as: gray, unstable risk; light pink, *Pf*PR$_{2–10}$ = 0–5%; midpink, *Pf*PR$_{2–10}$ = 5–40%; and dark red, *Pf*PR$_{2–10}$ = >40%. Pie chart size is representative of the relative number of migrants.

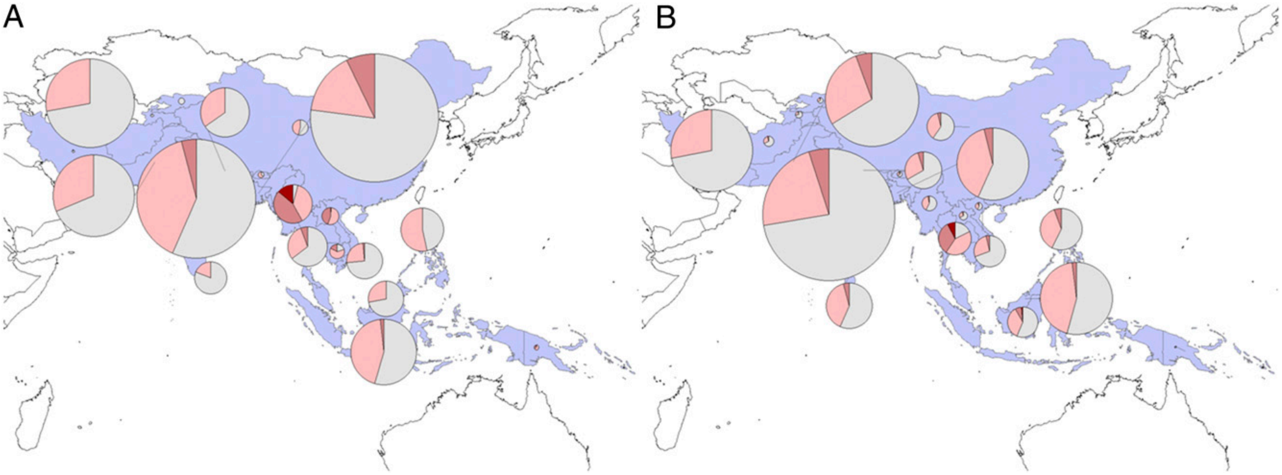

Fig. S5. CSE Asia: (*A*) outgoing migrants and (*B*) incoming migrants by endemicity class. Endemicity classes are colored as: gray, unstable risk; light pink, *Pf*$PR_{2–10}$ = 0–5%; midpink, *Pf*$PR_{2–10}$ = 5–40%; and dark red, *Pf*$PR_{2–10}$ = >40%. Pie chart size is representative of the relative number of migrants.