The Knanaya Ancestry Project:
Preliminary Genetic Analysis

August 28, 2007
E. Valaitis and L. Martin
DNA Tribes

About DNA Tribes

DNA Tribes is a private firm specializing in genetic ancestry analysis, including both geographical analysis of world populations and the comparison of individuals to living populations and world regions. DNA Tribes' proprietary analysis incorporates statistical techniques developed by Dr. Eduardas Valaitis, who received his Doctorate in Statistics from Yale University in 2005 and is now an Assistant Professor in the Department of Mathematics and Statistics at American University in Washington, D.C. Dr. Valaitis' background includes extensive work in multivariate analysis and classification, which involves identifying mathematical structure present within large and complex datasets. This expertise allows DNA Tribes to perform a uniquely detailed and comprehensive analysis of world populations to identify genetic structure on an objective mathematical basis.

About the Knanaya Ancestry Project

The Knanaya Ancestry Project aims to document the genetic affiliations of the Knanaya community in Kerala State, India. According to legend, in 345 AD, a Middle Eastern trader named Thomas of Cana (or Kana) migrated to Kerala, India accompanied by seven tribes with 72 families of Jewish Christians from Cana or Kana, in the Middle East.

The Knanaya Ancestry Project will evaluate this traditional narrative using modern genetics by comparing a Knanaya population sample to living populations around the world as well as major world regions. DNA collection from unrelated volunteer participants in the Knanaya community has been performed under the supervision of Mr. Cyriac Thannikary, who verified the Kana ethnicity of each participant. Testing for each participant includes a set of industry standard autosomal STR markers passed down from both parents and equally informative for both males and females. Although DNA samples from at least 100 people are required for a robust population analysis, this initial study evaluates a relatively small sample of Knanaya individuals as of August 2007 for comparison to Middle Eastern and other world populations.

Methods

First, the Knanaya population sample will be compared to individual ethnic groups in the DNA Tribes database of over 650 world populations using a distance measurement. This will provide a basic indication of which populations are individually most similar to Knanaya, which will be clarified by subsequent analyses. Next, this study will include a more sophisticated k-means analysis of DNA Tribes global population database including the Knanaya Ancestry Project population sample. K-means is a multivariate statistical method that assigns populations into distinct groups (clusters) based on their proximity in numerous data dimensions in relation to all other populations. Clustering is performed hierarchically as a series of progressively smaller divisions, from the large continental divisions to smaller world region groupings. This technique allows a multi-dimensional comparison of world populations to

identify genetic groups sharing a common pattern of genetic characteristics. Finally, an admixture analysis will be performed to evaluate the possibility of a limited genetic contribution from indigenous Kerala State populations.

Genetic Structure: Geographical or Historical

Using k-means, the genetic structure identified among world populations is primarily geographical. Genetic clusters (or groupings) generally correspond to geographical territories, often organized by natural features such as seas, deserts, mountains, and other land features. Most of these clusters are geographically discrete and can be described and mapped as genetic world regions (*see Figure 1 below*). Each of these genetic world regions is a territory sharing genetic characteristics shaped over the course of many generations.

Figure 1: Map of genetic world regions identified by DNA Tribes.

However, some populations with a history of recent migration and subsequent endogamy cluster along non-geographical lines, sharing closer connections with populations in their places or origin rather than present geographical neighbors. For instance, United States Caucasian populations cluster with Europeans, and some European Roma (Gypsy) populations cluster with Subcontinental India. In contrast, migrant populations with a history of intermarriage with local populations sometimes cluster with their present geographical neighbors. For instance, Spanish-speaking populations in many parts of Latin America cluster with Native Americans rather than Europeans, and Tatars living in Poland cluster with Europeans rather than Central Asians.

The geographical location of Knanaya communities of Kerala State, India predicts a genetic affiliation with autochthonous populations of Subcontinental India (termed the "Indian" world region). However, the Knanaya historical narrative describes a history of migration from the Middle East (within the territories of the "Asia Minor" genetic region) and subsequent endogamy in India. This predicts a genetic

affiliation with populations of Asia Minor rather than India. This study will evaluate each prediction (the first geographical, the second historical) as well as the third possibility of migration from Asia Minor with some local admixture in Kerala State.

Part One: Distance Comparison to Individual Ethnic Groups

To provide a basic, preliminary indication of which populations are most similar to Knanaya, the Knanaya population sample was compared to individual ethnic groups in the DNA Tribes database of over 650 world populations using a simple distance measurement of average absolute distance from each population sample (*see Table 1a below*). The populations most similar to Knanaya include many Middle Eastern ethnic groups, as well as nearby populations peripherally affiliated with this region such as Uyghur Turks and European Mediterraneans. Also included are some Indian populations and peripheral Indian-Middle Eastern populations such as Punjabis, who are primarily Subcontinental Indian but share some genetic characteristics with Asia Minor populations. This relationship to Indian populations could reflect limited admixture in Kerala State, which will be evaluated in Part Three of this study.

To further explore relationships between Knanaya and specifically Jewish populations, a distance comparison was performed to identify a ranking of Knanaya distance from Jewish population samples. In addition to the Israeli population sample ranked twelfth in distance from Knanaya (*see Table 1a below*), distance rankings from additional Jewish populations are listed after the top 20 most similar populations (*see Table 1b below*). This preliminary distance comparison to individual populations will be clarified by the more robust structural analysis in Part Two.

<table>
<tr><td colspan="2">**Table 1a**</td><td colspan="2">**Table 1b**</td></tr>
<tr><td>Rank</td><td>Population Name</td><td>Rank</td><td>Population Name</td></tr>
<tr><td>1</td><td>West Punjabi (Pakistan)</td><td>41</td><td>Sephardi Jewish</td></tr>
<tr><td>2</td><td>Dubai, United Arab Emirates</td><td>102</td><td>Israel</td></tr>
<tr><td>3</td><td>Uyghur (Western China)</td><td>214</td><td>Xuetes (Balearic Islands, Spain)</td></tr>
<tr><td>4</td><td>Pakistan</td><td>240</td><td>Ashkenazi (Budapest, Hungary)</td></tr>
<tr><td>5</td><td>United Arab Emirates</td><td>297</td><td>North African Jewish</td></tr>
<tr><td>6</td><td>Campania, Italy</td><td>354</td><td>Misrahi Jewish</td></tr>
<tr><td>7</td><td>Balearic Islands, Spain</td><td>359</td><td>Ashkenazi Jewish</td></tr>
<tr><td>8</td><td>Turkey</td><td></td><td></td></tr>
<tr><td>9</td><td>Turkey</td><td></td><td></td></tr>
<tr><td>10</td><td>Eastern Anatolia, Turkey</td><td></td><td></td></tr>
<tr><td>11</td><td>Kurdish (Northern Iraq)</td><td></td><td></td></tr>
<tr><td>12</td><td>Jewish (Israel)</td><td></td><td></td></tr>
<tr><td>13</td><td>Lingayat (Karnatake State, India)</td><td></td><td></td></tr>
<tr><td>14</td><td>Calabria, India</td><td></td><td></td></tr>
<tr><td>15</td><td>Turkey</td><td></td><td></td></tr>
<tr><td>16</td><td>East Indian (Canada)</td><td></td><td></td></tr>
<tr><td>17</td><td>Indian (Malaysia)</td><td></td><td></td></tr>
<tr><td>18</td><td>Southern Tunisia</td><td></td><td></td></tr>
<tr><td>19</td><td>Indian (Singapore)</td><td></td><td></td></tr>
<tr><td>20</td><td>Turkey</td><td></td><td></td></tr>
</table>

Table 1a (*left*) illustrates the top 20 world populations closest to Knanaya; Table 1b (*right*) illustrates the ranked distances of specifically Jewish reference populations. Please note that some groups are listed twice due to multiple reference samples.

Part Two: Hierarchical k-Means Analysis of Global Populations

To identify the genetic world region where Knanaya fit best, K-means analysis was performed on DNA Tribes' global population database including the Knanaya population sample. K-means is a more robust comparison fundamentally different from the basic distance metric performed in Part One, because it allows the simultaneous comparison Knanaya with all other world populations based on all dimensions of genetic information. K-means analysis identifies groups of populations that together share genetic characteristics distinguished from all other groupings. This analysis proceeded in three stages, from the most basic continental divisions to more detailed divisions corresponding to world regions, and recorded the cluster to which the Knanaya sample was assigned.

Stage 1 (Divisions in the World): Initial k-groupings divide all global populations into major continental groups. At k=5, five divisions are observed corresponding to five major continental groups: (1) Sub-Saharan African; (2) European and Middle Eastern; (3) Subcontinental India and Australian Aboriginal; (4) East Asian; and (5) Native American. At this stage, the Knanaya population sample clusters with group (2), European and Middle Eastern populations, rather than the group including all other Subcontinental Indian populations. This is significant, because several European Roma (Gypsy) population samples continue to cluster with Subcontinental India despite European admixture. Subsequent k-divisions confirm that Knanaya cluster at all times with European and Middle Eastern populations, and at no stage cluster with any Subcontinental Indian genetic grouping.

Stage 2 (Divisions in Europe and the Middle East): To further clarify the genetic relationship of Knanaya within European and Middle Eastern populations, hierarchical k-means analysis is performed within the European and Middle Eastern cluster observed in Stage 1. At k=2, a basic division is observed between (1) European and (2) Middle Eastern populations. At this stage, Knanaya cluster with group (2), including Middle Eastern populations. Subsequent k-divisions (k=3 through k=5) confirm that Knanaya cluster with Middle Eastern populations at all times and at no stage with a European genetic grouping.

Stage 3 (Divisions in the Middle East): The third stage of the hierarchical clustering then proceeded to perform k-means within the Middle Eastern grouping observed in Stage 2. At k=3, three groupings are observed within this Middle East: (1) North African; (2) Asia Minor; and (3) Arabian. At this k-stage, Knanaya Ancestry Project participants cluster with Asia Minor group (2) (*see map in Figure 1*), which also includes Turkic, Persian, and Misrahi Jewish populations.

At k=4, four groupings are observed within this Middle Eastern cluster: (1) North African; (2) Southern Arabian; (3) Asia Minor; and (4) Northern Arabian and Jewish populations. At this k-stage, Knanaya Ancestry Project participants cluster with the Southern Arabian group (2). Note: Some Southern Arabian populations (such as the United Arab Emirates) include resident individuals from Subcontinental India. If Knanaya populations include a component of Indian ancestry from occasional gene exchange with local Subcontinental Indian populations, this could be responsible for observed clustering with Southern Arabia at this stage.

At k=5, Knanaya are comprise their own singleton cluster. These results indicate Knanaya are a primarily Middle Eastern population that possesses genetic characteristics less typical of other populations within this region. This could reflect factors including historical endogamy, a limited amount of admixture in Kerala State, or the small Knanaya sample size used for this comparison.

Part Three: Estimating Middle Eastern and Subcontinental Indian Genetic Contributions

K-means analysis in Part Two identifies Knanaya as primarily affiliated with Middle Eastern populations. Although this indicates that autochthonous origins in Subcontinental India are unlikely, metric proximity to Indian populations identified in Part One suggests the possibility of limited gene exchange with local Indian populations since historical settlement in Kerala State is not excluded. To estimate the percentages of genetic contribution from (1) Asia Minor and (2) Subcontinental Indian populations, an admixture analysis is performed using available Knanaya population data. This analysis proceeded by simulating populations of all proportions of Asia Minor and Subcontinental Indian admixture (from fully Asia Minor to fully Subcontinental Indian, along with all intermediate proportions), and then identifying which proportion of admixture most closely approximates observed Knanaya genetic characteristics.

This admixture analysis indicates that a population of 54.3% genetic contribution from Asia Minor, and 45.7% contribution from Subcontinental India would be the least distant from the Knanaya population sample. This estimate that a possibility of limited intermarriage with local populations of Kerala State cannot be excluded. However, this percentage estimate could be influenced by the small number of sampled Knanaya individuals or local Knanaya genetic characteristics due to historical endogamy, and might not fully represent the larger Knanaya community.

Summary and Conclusion of Preliminary Study

The preliminary distance comparison to individual ethnic groups in Part One indicates Knanaya are most similar to Middle Eastern populations as well as some Indian populations. The distance comparison to Jewish population samples indicated that among Jewish reference populations, Knanaya are least distant from Israeli and Sephardi populations (*see Tables 1a and 1b in Part One above*).

The more sophisticated multidimensional k-means clustering exercise in Part Two indicates that the population sample of Knanaya Ancestry Project participants is primarily affiliated with Middle Eastern populations, rather than with Subcontinental India. Within the Middle East, Knanaya comprise their own singleton cluster distinct from other Middle Eastern groupings. This could reflect historical endogamy, which would have the effect of fostering private genetic characteristics distinguishing Knanaya from source populations in the Middle East. Alternatively, this genetic outlier status could reflect the small size of the Knanaya population sample available at this time, or a limited genetic contribution from local populations in Kerala State. The collection of additional Knanaya population data would make this inference more precise, but might not alter this observed distinctiveness.

Admixture analysis in Part Three estimates that some genetic contribution from local Kerala State populations cannot be excluded. However, estimated proportions of Middle Eastern and Subcontinental Indian populations could reflect the small available Knanaya population sample and might not be fully representative of the Knanaya community.

Results of this preliminary study indicate the Knanaya are genetically affiliated with Middle Eastern populations, confirming the traditional historical narrative of migration from the Middle East. Results also indicate that Knanaya stand out as a genetically unique population among Middle Eastern genetic groups. This might reflect historical endogamy, a limited genetic contribution from local Kerala State

populations, or the small pool of Kana individuals available at the time of this study. Further data collection could allow a clarification of these preliminary observations.

For more information on Knanaya Ancestry Project, please contact:

Cyriac Thannikary
137 Minuteman Circle
Allentown, NJ 08501
Phone: 609-208-9280
E-mail: Thannikary@aol.com