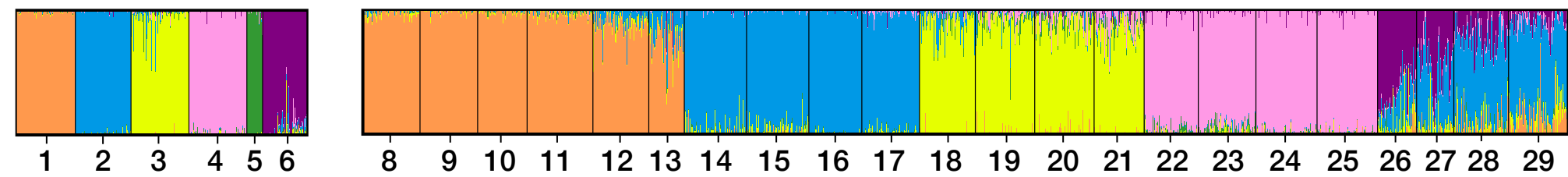


Supplementary Fig. S2.1. Bio-geographical ancestry analysis based on the six reference populations (1-6, summarised in Fig. 1) plus unadmixed (8-11, 14-25) and admixed (12-13, 26-29) populations available from 1000 Genomes 30X high coverage data (1KG-30). **(A)** STRUCTURE results of cluster membership proportions at K=6, with all non-reference populations classified against the reference dataset 1-6. For the STRUCTURE analysis, the inferred clusters are colored so the major proportion of each reference population matches the righthand legend. **(B)** Three dimensional MDS analysis showing coordinates 1 and 3 (top plot), 1 and 2 (bottom). **(C)** Neighbor Joining Tree (NJT) analysis. For the MDS and NJT plots, populations are colored according to the righthand legend. Note, AMR reference set 6 includes 18 PEL without co-ancestry and the remaining 67 admixed PEL are in set 26.

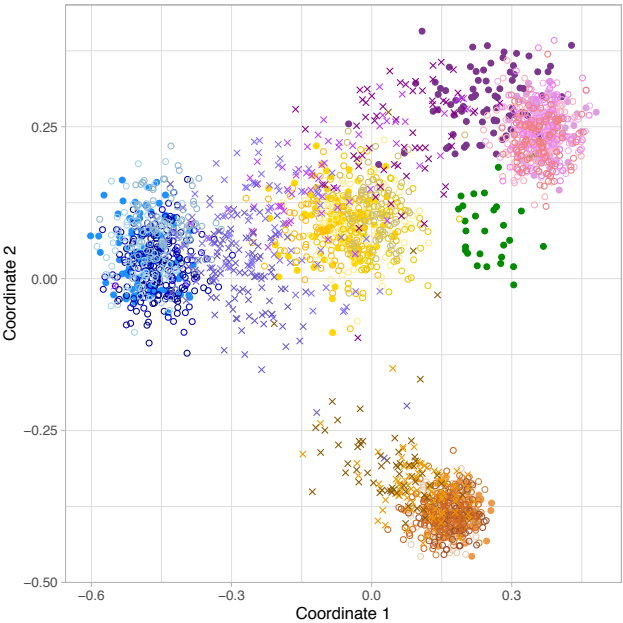
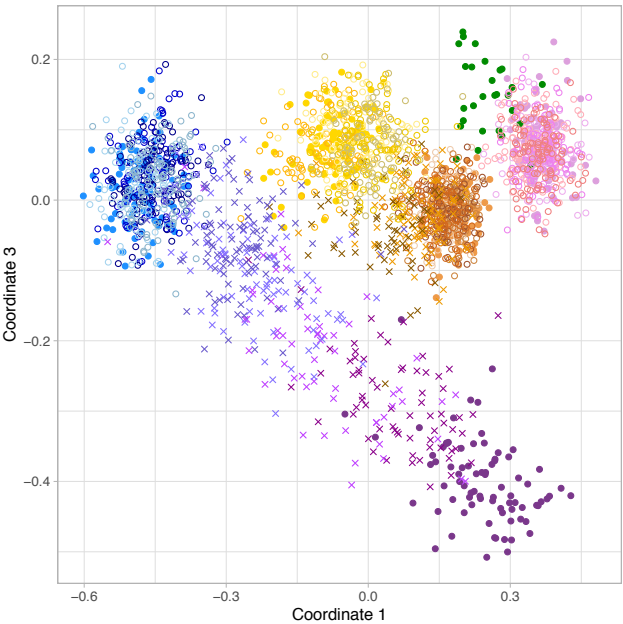
A



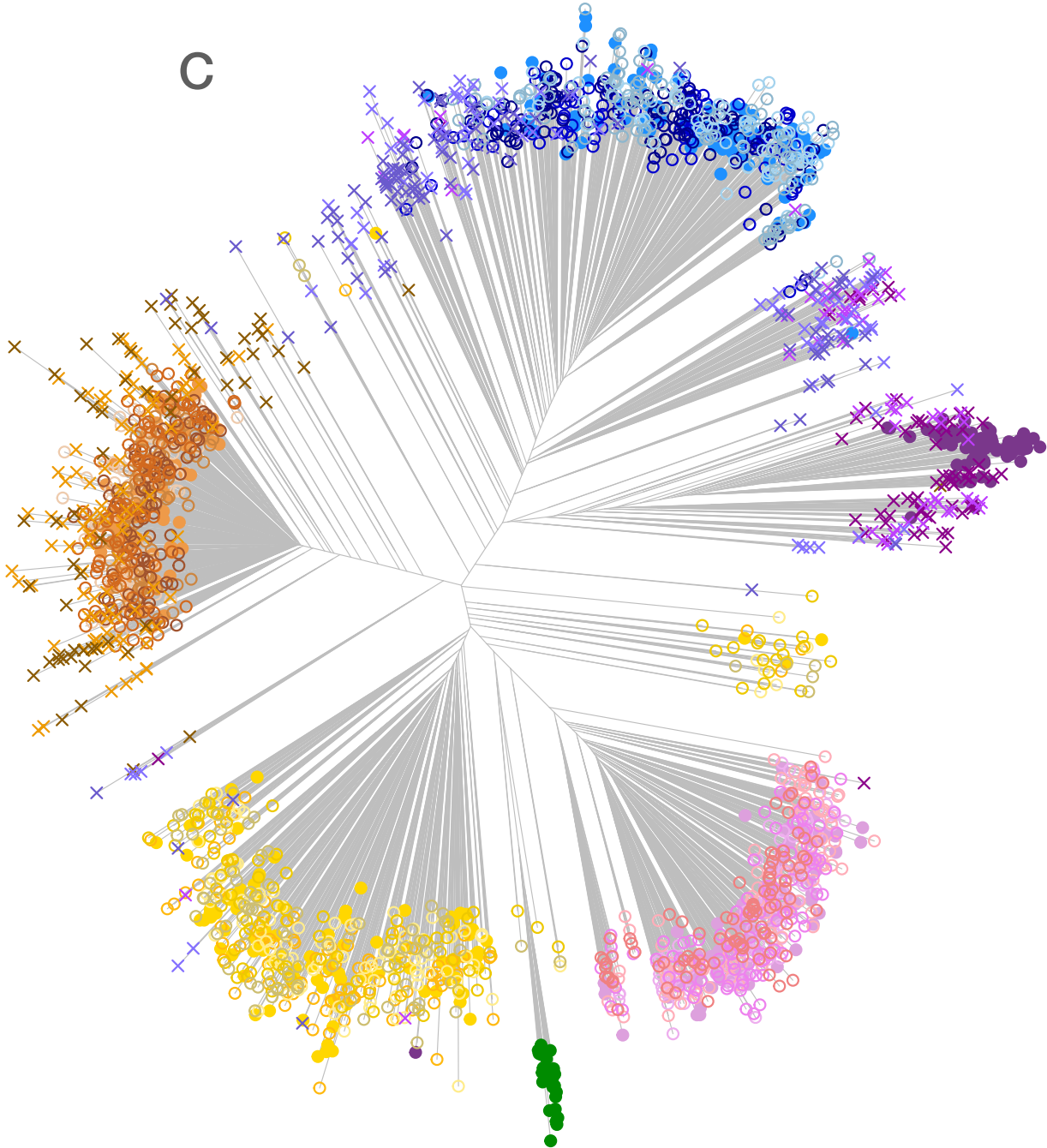
Reference populations

- | | | | | | |
|---|---|-----|---|---|-----|
| 1 | ● | AFR | 4 | ● | EAS |
| 2 | ● | EUR | 5 | ● | OCE |
| 3 | ● | SAS | 6 | ● | AMR |

B



C



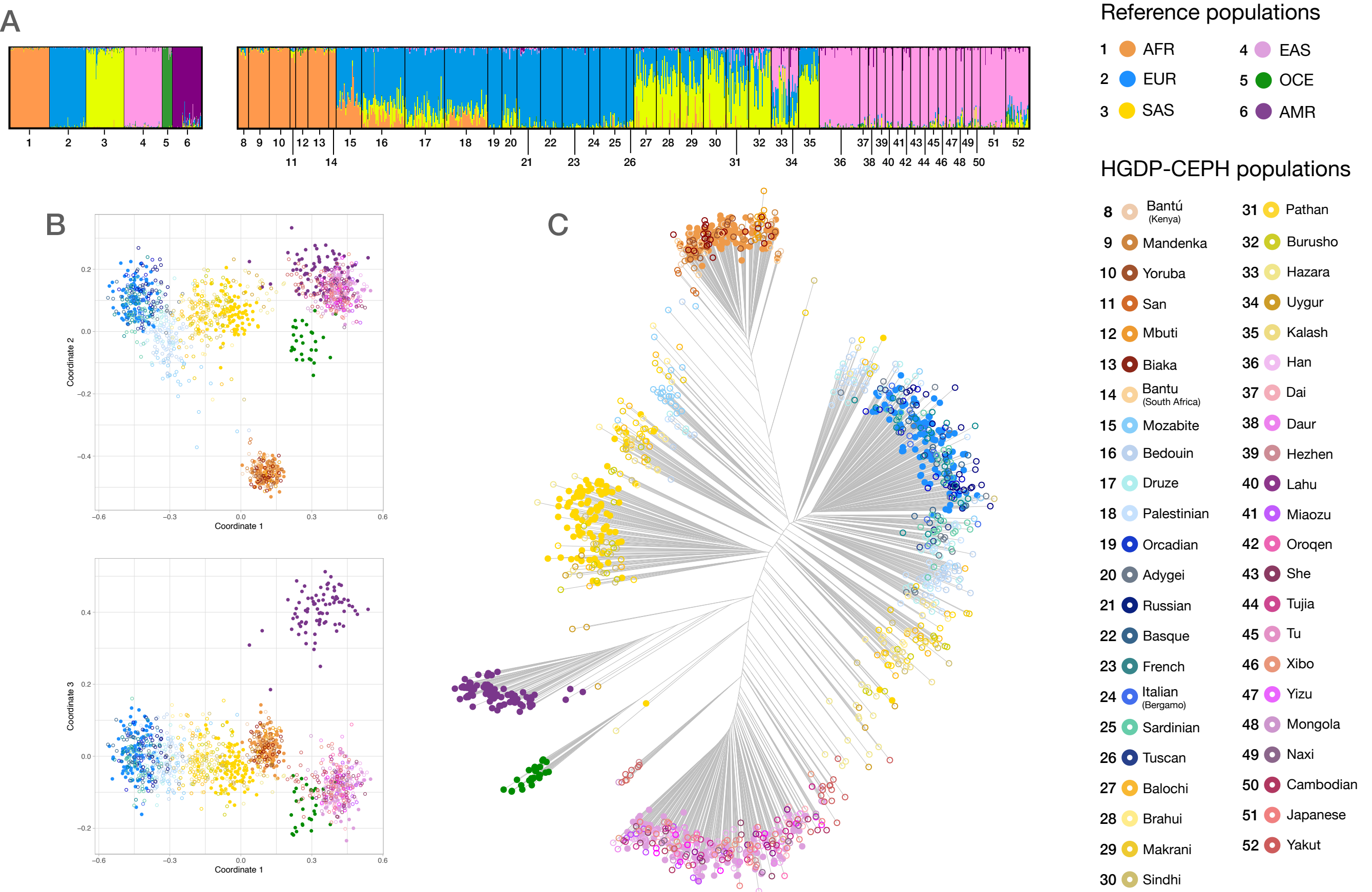
1KG-30 Unadmixed

- | | | | | | |
|----|---|-----|----|---|-----|
| 8 | ○ | LWK | 18 | ○ | PJL |
| 9 | ○ | ESN | 19 | ○ | ITU |
| 10 | ○ | MSL | 20 | ○ | STU |
| 11 | ○ | GWD | 21 | ○ | BEB |
| 14 | ○ | TSI | 22 | ○ | CDX |
| 15 | ○ | IBS | 23 | ○ | KHV |
| 16 | ○ | GBR | 24 | ○ | CHS |
| 17 | ○ | FIN | 25 | ○ | JPT |

1KG-30 Admixed

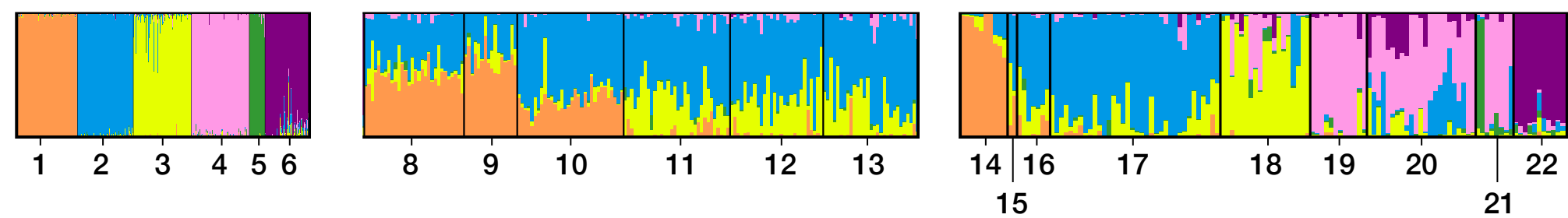
- | | | | | | |
|----|---|------------------|----|---|-----|
| 12 | × | ACB | 27 | × | MXL |
| 13 | × | ASW | 28 | × | CLM |
| 26 | × | PEL
(admixed) | 29 | × | PUR |

Supplementary Fig. S2.2. Bio-geographical ancestry analysis based on the six reference populations presented in Fig. 1, plus populations available from HGDP-CEPH Whole Genome Sequencing Data. **(A)** STRUCTURE results of ancestry proportions at K=6, non-reference populations were classified against the lefthand reference dataset 1-6. For the STRUCTURE analysis, the inferred clusters are colored so the major proportion of each reference population (1-6) matches the righthand legend. **(B)** Three dimensional MDS analysis showing coordinates 1 and 2 (up) and 1 and 3 (down). **(C)** Neighbor Joining Tree (NJT) analysis. For the MDS and NJT plots, populations are colored according to the righthand legend.



Supplementary Fig. S2.3. Bio-geographical ancestry analysis based on the six reference populations presented in Fig. 1, plus populations available from in-house genotyping with the VISAGE BT tool and SGDP data. **(A)** STRUCTURE results of cluster membership proportions at K=6, non-reference populations were classified against the lefthand reference dataset 1-6. For the STRUCTURE analysis, the inferred clusters are colored so the major proportion of each reference population matches the righthand legend. **(B)** Three dimensional MDS analysis showing coordinates 1 and 2 (top), 1 and 3 (bottom). **(C)** Neighbor Joining Tree (NJT) analysis. For the MDS and NJT plots, populations are colored according to the righthand legend.

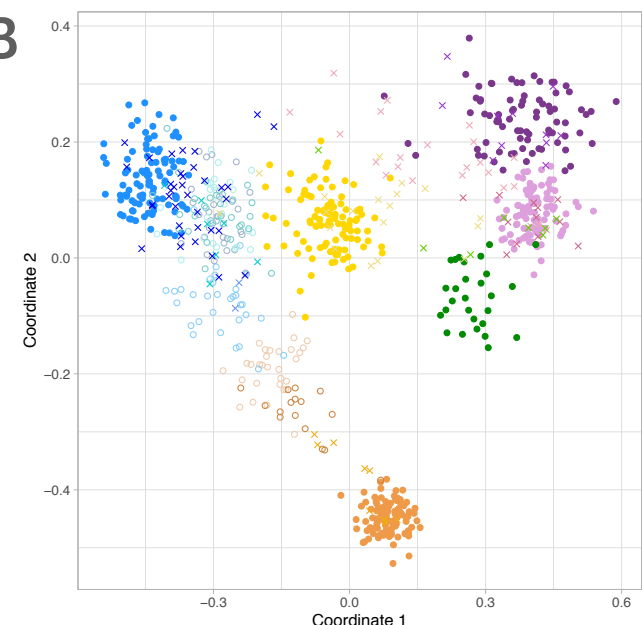
A



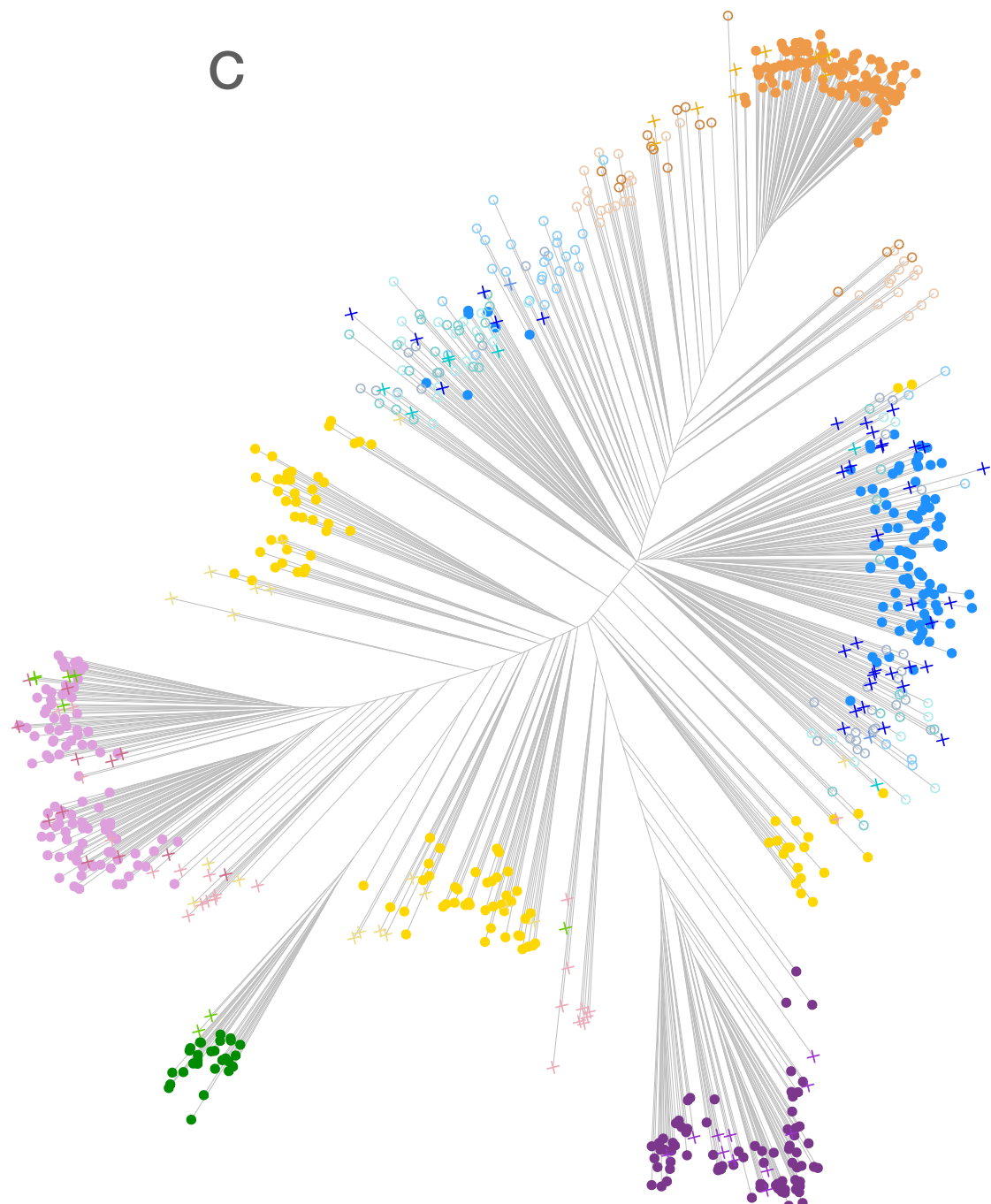
Reference populations

- | | | | | | |
|---|---|-----|---|---|-----|
| 1 | ● | AFR | 4 | ● | EAS |
| 2 | ● | EUR | 5 | ● | OCE |
| 3 | ● | SAS | 6 | ● | AMR |

B



C



VISAGE in-house pops

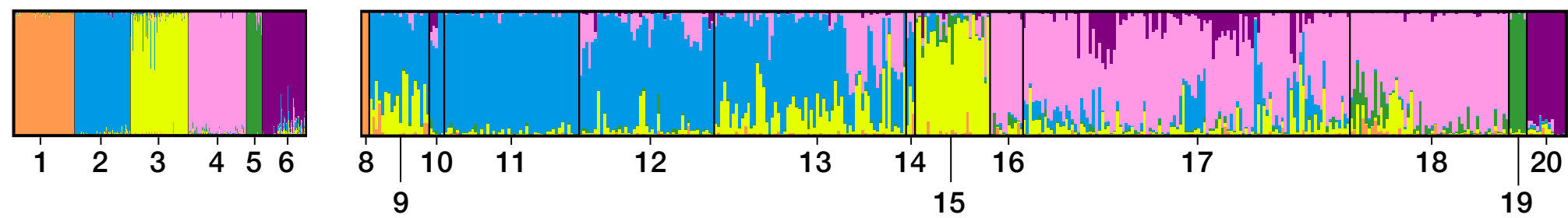
- | | | |
|----|---|-------------------------|
| 8 | ○ | Eritrea |
| 9 | ○ | Somalia |
| 10 | ○ | Morocco |
| 11 | ○ | Iraq (Central region) |
| 12 | ○ | Iraq (Kurdistan region) |
| 13 | ○ | Turkey |

SGDP population groups

- | | | |
|----|---|----------------|
| 14 | × | Africa |
| 15 | × | North Africa |
| 16 | × | Middle East |
| 17 | × | Europe |
| 18 | × | South Asia |
| 19 | × | East Asia |
| 20 | × | Northeast Asia |
| 21 | × | Oceania |
| 22 | × | America |

Supplementary Fig. S2.4. Bio-geographical ancestry analysis based on the six reference populations presented in Fig. 1, plus populations available from EGDP data, with ~85% (95/115 SNPs) profile completeness. **(A)** STRUCTURE results of ancestry proportions at K=6, non-reference populations were classified against the reference dataset. For the STRUCTURE analysis, the inferred clusters are colored so the major proportion of each reference population (1-6) matches the righthand legend. **(B)** Three dimensional MDS analysis showing coordinates 1 and 2 (top plot), 1 and 3 (bottom). **(C)** Neighbor Joining Tree (NJT) analysis. For the MDS and NJT plots, populations are colored according to the righthand legend and the analysis is based on the 100 SNPs present in the EGDP data.

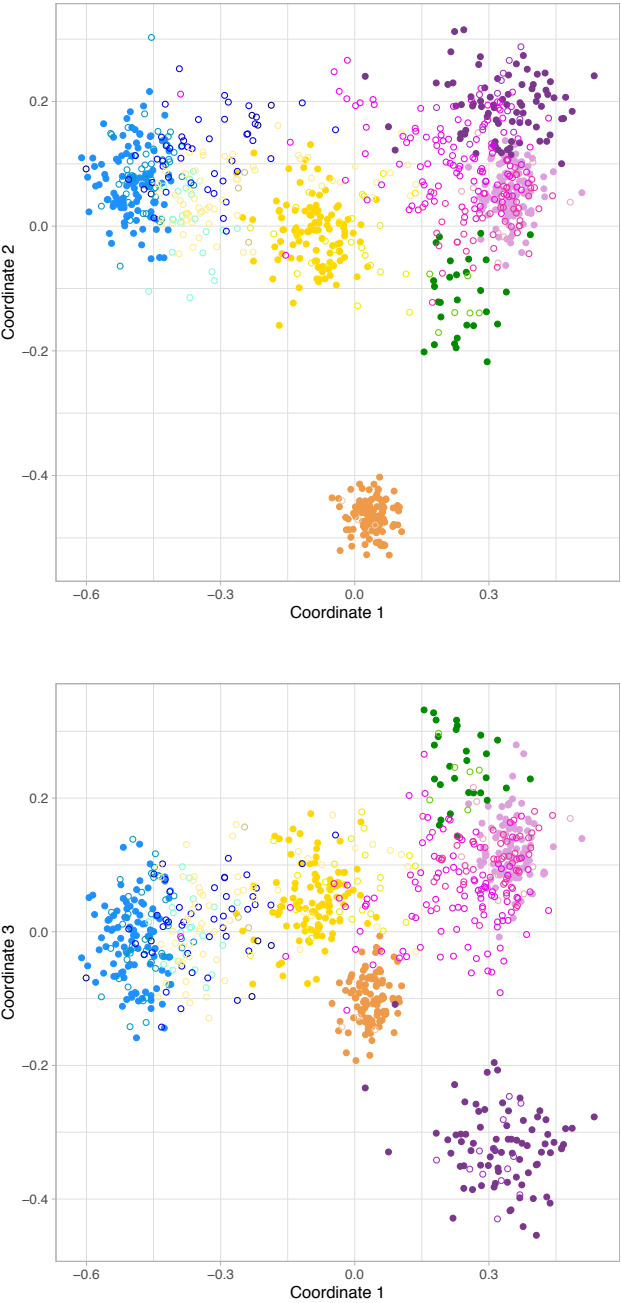
A



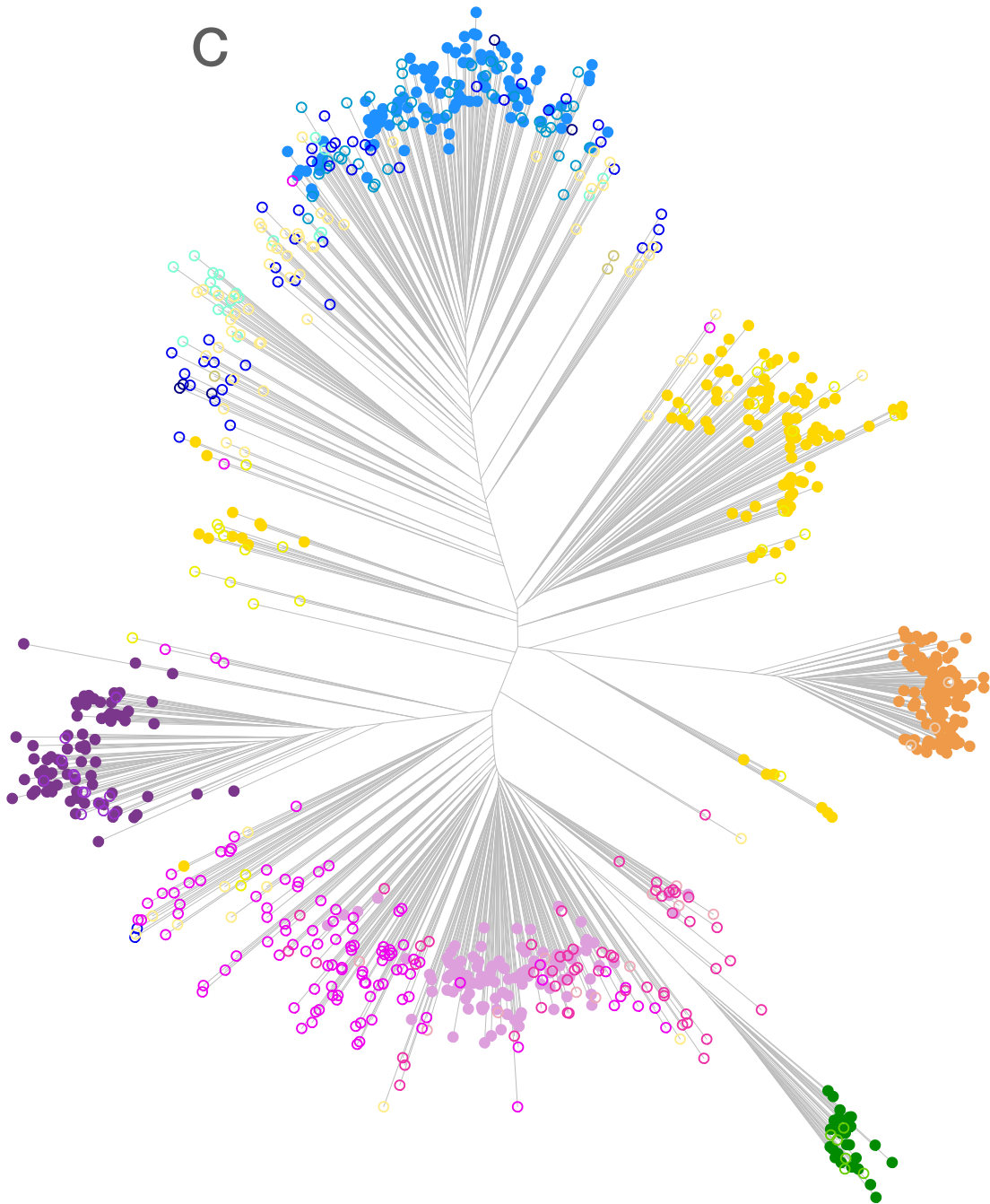
Reference populations

- | | | | | | |
|---|---|-----|---|---|-----|
| 1 | ● | AFR | 4 | ● | EAS |
| 2 | ● | EUR | 5 | ● | OCE |
| 3 | ● | SAS | 6 | ● | AMR |

B



C



EGDP populations

- | | | |
|----|---|--------------------|
| 8 | ● | Sub-Saharan Africa |
| 9 | ● | Middle East |
| 10 | ● | North Europe |
| 11 | ● | Europe |
| 12 | ● | Eastern Europe |
| 13 | ● | Central Asia |
| 14 | ● | Roma |
| 15 | ● | South Asia |
| 16 | ● | East Asia |
| 17 | ● | Northeast Asia |
| 18 | ● | Southeast Asia |
| 19 | ● | Papua New Guinea |
| 20 | ● | America |