

Response to: A commentary on “A Vietnamese human genetic variation database”

Vinh S. Le¹, Kien T. Tran¹, and Liem T. Nguyen¹

¹Vinmec Research Institute of Stem Cell and Gene Technology, 458 Minh Khai, Hai Ba Trung, 10000 Hanoi, Vietnam.

Correspondence

Vinh Sy Le, Vinmec Research Institute of Stem Cell and Gene Technology, 458 Minh Khai, Hai Ba Trung, Ha Noi, Vietnam

Email: vinhls@vnu.edu.vn

Liem Thanh Nguyen, Vinmec Research Institute of Stem Cell and Gene Technology, 458 Minh Khai, Hai Ba Trung, Ha Noi, Vietnam

Email: v.liemnt@vinmec.com

This letter is a response to the commentary by Jonson & Do (Johnson and Do 2020) on our paper, entitled “A Vietnamese human genetic variation database” (Vinh et al. 2019). The commentators concerned about two issues: Firstly, the relation of Southeast Asian (SEA) and East Asian (EA) groups to African and European groups; Secondly, the history of migration and settlement in Southeast Asia. Our responses will clarify both concerns from the commentators.

The first concern is about our statement “The tree structures show that SEA populations are closer to the YRI and CEU than EA populations” inferred from the phylogenetic tree in Figure 4 of Vinh et al. 2019 (or Figure 1 in this paper). The commentators concerned that “the relative positions of SEA and EA in the figure, with SEA closer to the YRI and CEU, is an artifact of the presentation, and must not be understood as implying a closer relationship to Africans and European”. Our statement was based on the full tree structure containing all SEA and EA individual populations, not from a quartet-tree of only for leaves (i.e., YRI, CEU, SEA, and EA). The full tree structure obviously shows that the topological distance (i.e., the number of branches between two nodes) between an individual SEA population to the YRI is smaller than that between an individual EA population to the YRI. For example, the topological distance $d(\text{MY}, \text{YRI})$ between Malay Malaysia (MY) and YRI is 3 and smaller than that between Japanese (JPT) and YRI (i.e. $d(\text{JPT}, \text{YRI}) = 10$). Note that the topological distances among populations do not depend on the presentation/visualization of the tree structure.

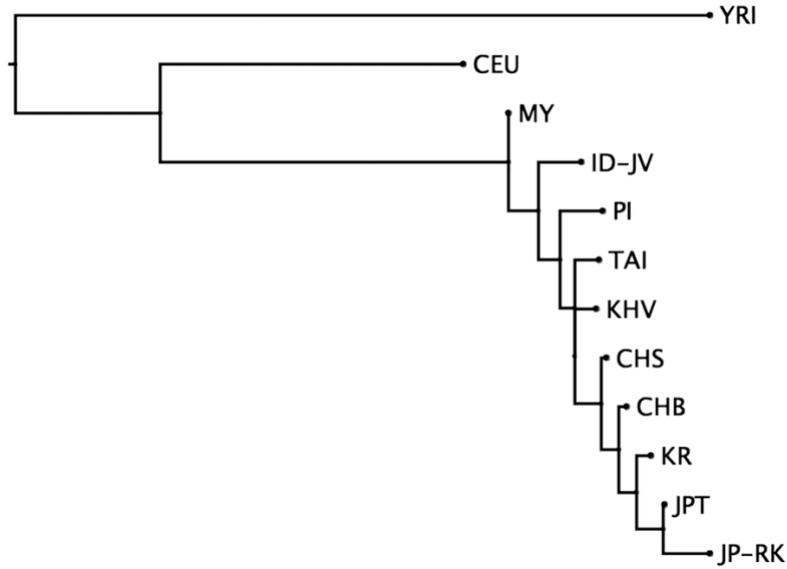


Figure 1: The phylogenetic tree structures of 12 populations as illustrated in Figure 4 of (Vinh et al. 2019).

In this letter, we additionally provide the F_{ST} distances among populations (Table 1) which we used to build the phylogenetic tree by the neighbor-joining method. A smaller F_{ST} distance value between the two populations indicates a closer genetic relationship between them. Hence, the F_{ST} distances show that the genetic distances between the SEA populations and the YRI are smaller than that between the EA populations and the YRI.

Table 1: The F_{ST} distances among individual populations

	YRI	CEU	MY	ID-JV	PI	TAI	KHV	CHS	CHB	KR	JPT	JP-RK
YRI	0	0.166	0.169	0.185	0.189	0.19	0.188	0.19	0.191	0.193	0.193	0.192
CEU	0.166	0	0.088	0.108	0.112	0.112	0.11	0.113	0.113	0.116	0.114	0.115
MY	0.169	0.088	0	0.006	0.01	0.01	0.012	0.013	0.018	0.022	0.024	0.029
ID-JV	0.185	0.108	0.006	0	0.013	0.01	0.011	0.016	0.021	0.026	0.028	0.034
PI	0.189	0.112	0.01	0.013	0	0.012	0.012	0.011	0.016	0.019	0.022	0.029
TAI	0.19	0.112	0.01	0.01	0.012	0	0.005	0.006	0.01	0.015	0.018	0.026
KHV	0.188	0.11	0.012	0.011	0.012	0.005	0	0.006	0.009	0.014	0.017	0.026
CHS	0.19	0.113	0.013	0.016	0.011	0.006	0.006	0	0.002	0.006	0.011	0.02
CHB	0.191	0.113	0.018	0.021	0.016	0.01	0.009	0.002	0	0.003	0.007	0.018
KR	0.193	0.116	0.022	0.026	0.019	0.015	0.014	0.006	0.003	0	0.004	0.015
JPT	0.193	0.114	0.024	0.028	0.022	0.018	0.017	0.011	0.007	0.004	0	0.007
JP-RK	0.192	0.115	0.029	0.034	0.029	0.026	0.026	0.02	0.018	0.015	0.007	0

Perhaps, the misinterpretation of the commentators was due to their consideration of a quartet-tree $T((YRI, (CEU, (SEA, EA))))$ with only four leaves representing for four population groups, i.e., YRI, CEU, SEA, and EA. The structure of quartet-tree $T((YRI, (CEU, (SEA, EA))))$ should not be used compare the relationships among SEA, EA and YRI. Readers should interpret the topological relationships among populations using the full tree structure containing all individual populations.

The second concern from the commentators is the history of migration and settlement in Southeast Asia that we summarized based on the findings from two papers (Lipson et al. 2018; McColl et al. 2018), i.e., “the findings agree with our ancestral population analyses that the present-day SEA populations were mainly derived from the SEA ancestries and partly from the EA ancestries”. Our statement simply referred to the main findings from (Lipson et al. 2018; McColl et al. 2018) that were clearly stated in the abstract of McColl et al. 2018, i.e., "Both Hòabínhian hunter-gatherers and East Asian farmers contributed to current Southeast Asian diversity, with further migrations affecting island SEA and Vietnam".

The contribution of the EA ancestries to the present-day SEA populations can be perceived from the figure S11 in McColl et al. 2018. This figure describes the contribution of ancestral populations to the present-day populations with different K values ranged from 2 to 13. At $K \geq 6$, the EA ancestries (illustrated in dark green color) and the SEA ancestries (illustrated in pink color) are separated. The figure clearly shows that the EA ancestries partly contributed to both present-day island SEA (ISEA) and mainland SEA (MSEA) populations. Specifically, the present-day ISEA and MSEA populations were mainly derived from the SEA ancestries and partly from the EA ancestries. Note that the majority contribution of the SEA ancestries to the present-day SEA populations was also obviously observed from the Figure 5 in our paper (Vinh et al. 2019).

We understand that the human history of the SEA populations is complicated. Therefore, we encourage other groups to have further investigations and collaborations in this interesting research area. Finally, our database has been publicly available on <https://genomes.vn/> for academic purposes. We are willing to collaborate with other groups in using our dataset for solving interesting problems.

DISCLOSURE STATEMENT

The authors declare no conflict of interest.

REFERENCES

- Johnson, Adam, and Linh Do. 2020. "A Commentary on "A Vietnamese Human Genetic Variation Database"." *Human Mutation*, humu-2019-0590.R1.
- Lipson, Mark, Olivia Cheronet, Swapan Mallick, Nadin Rohland, Marc Oxenham, Michael Pietrusewsky, Thomas Oliver Pryce, et al. 2018. "Ancient Genomes Document Multiple Waves of Migration in Southeast Asian Prehistory." *Science (New York, N.Y.)* 361 (6397): 92–95. <https://doi.org/10.1126/science.aat3188>.
- McColl, Hugh, Fernando Racimo, Lasse Vinner, Fabrice Demeter, Takashi Gakuhari, J. Víctor Moreno-Mayar, George Van Driem, et al. 2018. "The Prehistoric Peopling of Southeast Asia." *Science* 361 (6397): 88–92. <https://doi.org/10.1126/science.aat3628>.
- Vinh, Le Sy, Kien Tran, Hoa Bui, Huong Le, Canh Nguyen, Duong Do, Ha Ly, Linh Pham, Lan Dao, and Liem Nguyen. 2019. "A Vietnamese Human Genetic Variation Database." *Hum*, no. 40: 1664–1675. <https://doi.org/doi:10.1002/humu.23835>.