An MPI-based parallel algorithm to study tribal fission and founders effect

Wes Brewer
Computational Solutions LLC
wes@computational.io

Abstract—In this paper we present parallel computational model for studying tribal dynamics in population genetics based on the Message Passing Interface (MPI). We describe three different fission models including: fission doubling, radial divergence, and competition fission. We apply this model to study to phenomenon of founder effect, demonstrated using both fission doubling and radial divergence.

Keywords — MPI, founder effect, group selection

I. Background

In this paper, we look into the effect of population substructure on natural selection. To study these effects we added the ability within Mendel's Accountant, a biologically-realistic population genetics simulator, to handle population substructure with tribal competition. We also attempt to model the evolution of altruistic traits. This allows us to hopefully shed new light on some longstanding theoretical questions regarding the validity of group selection as a means to account for social behavior.

II. Implementation

Mendel's Accountant is a forward-time population genetics simulation program, which models genetic change over time [1]. Mendel's Accountant primarily focuses on biological realism in studying natural selection of human populations, but has also been extended to work for viruses (e.g. [2]), as well as bacterial genomes. Mendel's Accountant has also been used to study a number of problems such as genetic load [3], nearly neutral mutations accumulation, and waiting times for fixation of multiple co-dependent mutations [4].

A. Population substructure

Migration of an individual from one tribe to another is modeled by transferring that individual's genetic information from one process to another using the Message Passing Interface (MPI). In general, each tribe is assigned to a separate processor, although with MPI it is possible to assign multiple tribes/processes to each processor. Communication of the genetic information of a migrating individual is performed asynchronously via standard non-blocking MPI Isend and Irecv calls. For each migrating individual, four types of information are communicated to the destination process: (1) the list of integers encoding the tracked deleterious mutations, (2) the list of integers encoding the tracked favorable mutations, (3) the list of fitnesses for each linkage block, and (4) the list of the total number of mutations in each linkage block. Before communication is performed, the four lists are gathered together from each of the randomly selected migrating individuals and packed into communication buffers. Data in the buffers are then transmitted to the appropriate destination. Mendel supports three types of migration models as listed here and illustrated in Fig. 1:

1) **Round robin** – this is a ring pass structure, where each tribe passes individuals to the tribe on its right, tribe (i+1)%num_tribes, and receives individuals from the tribe on its left, tribe (i-1)%num_tribes
2) **Stepping stone** – all tribes are organized in a ring, and each tribe both sends and receives individuals from its neighbor to its right, and the neighbor to its left.
3) **Island model** – every tribe sends and receives individuals from every other tribe.

B. Tribe structure

If we consider the population of a single tribe, we can compute a conservation of individuals to develop an expression that governs the population in a given tribe and its functional dependence on the other relevant variables as follows:

$$N_{i+1} = N_i + B_i - D_i + I_i - E_i \tag{1}$$

In this equation, for a given generation i, N_i is the number of individuals in a population. B_i is the number of new births that occur, which can be defined as:

$$B_{i+1} = \left(\frac{N_i}{2}\right) F_i - S_i$$

Where F is the total fertility rate, and S_i is the surplus individuals selected away. If the population is static, $B_i - D_i = 0,$ S can be defined as:

$$S_{i+1} = \left(\frac{N_i}{2}\right) F_i - N_i = N_i \left(\frac{F_i}{2} - 1\right)$$

For example, for a population size N = 1000 and F = 6, S would be 2000. D_i is the number of deaths and can be defined by:

$$D_{i+1} = (1 + f_{RD}) N_i$$

There are two components to the death rate: (1) In Mendel's Accountant, all the individuals are replaced by their successors thereby assuming they all die (N_i), and (2) there are an additional $f_{RD} N_i$ number of individuals that die a random death before mating occurs. Here, f_{RD} is the fraction of individuals that die a random death. I is the immigrants coming into the population from another tribe, and E is the number of emigrants leaving the population going to another tribe. In the special case that $I_i - E_i = 0$, for every immigrant coming into the tribe from another population, there is a corresponding emigrant leaving the tribe going to another tribe. Thus, there is no net increase or decrease of the individual tribes due to tribal migration. Currently, Mendel's Accountant always works in this way.

For a tribe to grow, the birth rate must be greater than the death rate. This is currently handled simply by changing the reproductive rate during the times of growth. In the current implementation, we essentially turn off selection during the growth process so that no individuals are selected away during the growth process.

C. *Fission model*

Implementing fission using MPI is difficult primarily due to the difficulties of doing dynamic process management in MPI. Talking about MPI_SPAWN, MPI's method of dynamic process create, Jack Dongarra states [5]: *"The prospects for dynamic process management… are much more difficult to implement, and are impossible to layer on top of existing implementations."*

After spending some time grappling with the difficulties of implementing dynamic process management, we opted to simplify the problem by using static processes, which have a rather well developed infrastructure within MPI. To handle dynamic population creation using static population, we created two different types of tribes that are identified with a *state* flag variable. Each tribe can either be in a LIVE state or a ZOMBIE state. While this is computationally more expensive than using dynamic tribes, it offers several benefits: (1) simpler, more robust implementation, and (2) allows the tribes to be pre-initialized and remain synchronized in time to the other tribes (i.e. every tribe is synchronized to the same generation in biological time, and does not require any burn-in time to start up), (3) can be implemented using multiple MPI communicator groups.

We have implemented three types of fission models in Mendel, two of which are illustrated in Fig. 2, so called *fission doubling* and *radial divergence*. A third fission

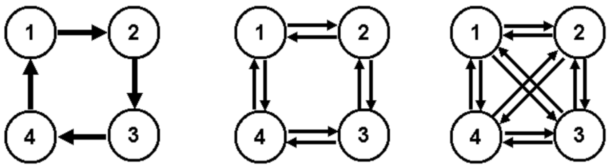

Figure 1. Migration models in Mendel:
(a) Round-robin, (b) Stepping Stone, (c) Island

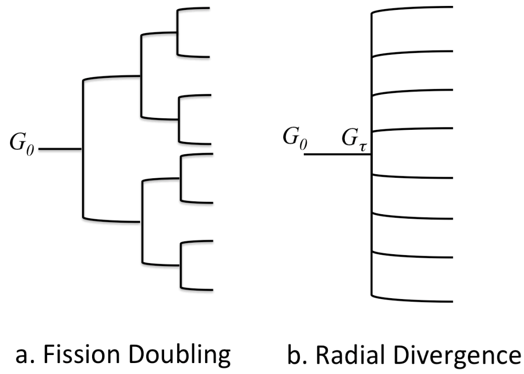

Figure 2. Tribal fission models in Mendel

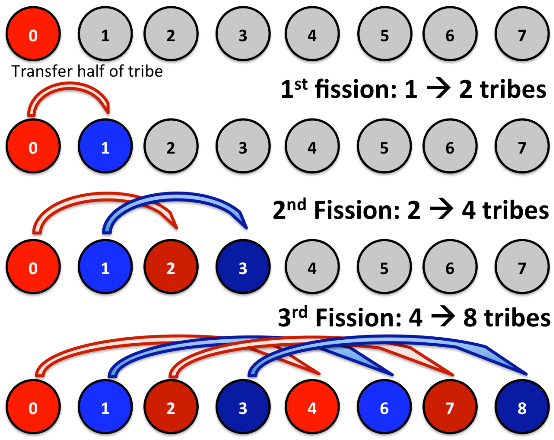

Figure 3. Destination tribe selection for fission doubling

model is used for group competition, which simply sends half of the winning tribes' individuals to replace individuals in the losing tribe. Here we will further explain fission doubling and radial divergence fissioning by using an example of eight tribes.

1) **Fission doubling**. Fission doubling essentially splits a tribe in two when a critical population size threshold τ is reached. In this case we choose a destination tribe using the following function using the following pseudocode:

```
if (tribe_state == LIVE) then
 partner = mod(tribe+num_tribes, 2*num_tribes)
```

`partner` here represents the partnering tribe, and `mod` is the modulo function. The results of this selection process are shown in Fig. 3. We also determine who is the sender and receiver simply by the logical test:

```
myid < partner : sender
myid > partner : receiver
```

For eight tribes, this results in the following states, which also corresponds with Fig. 2a. After the first fission event:

Tribe	Pre-fission state	Post-fission state	Send to/ Receive from	Partner
0	Live	Live	Send	1
1	Zombie	Live	Receive	0
2	Zombie	Zombie	Send	5
3	Zombie	Zombie	Send	6
4	Zombie	Zombie	Send	7
5	Zombie	Zombie	Receive	2
6	Zombie	Zombie	Receive	3
7	Zombie	Zombie	Receive	4

After the second fission event:

Tribe	Post-fission state	Send to/ Receive from	Partner
0	Live	Send	2
1	Live	Send	3
2	Live	Receive	0
3	Live	Receive	1
4	Zombie	Send	6
5	Zombie	Send	7
6	Zombie	Receive	4
7	Zombie	Receive	5

After the third fission event:

Tribe	Post-fission state	Send to/ Receive from	Partner
0	Live	Send	4
1	Live	Send	5
2	Live	Send	6
3	Live	Send	7
4	Live	Receive	0
5	Live	Receive	1
6	Live	Receive	2
7	Live	Receive	3

2) **Radial divergence** – radial divergence is used for fissioning more than two tribes in a single event. This could be useful to model a tower of Babel type event, where a single tribe was split into approximately 16 tribes in a single event. In this case either generation $G_{\mathcal{T}}$ or population size threshold may be used to trigger the fission event. Fig. 2b shows an example of radial fission for eight tribes. At the fission event, tribe 0 sends `NP/num_tribes` individuals to every tribe except itself. So, after the fission event, the status looks like.

Tribe	Pre-fission	Post-fission	Send to / Receive from	Partner
0	Live	Live	Send	1-7
1	Zombie	Live	Receive	0
2	Zombie	Live	Receive	0
3	Zombie	Live	Receive	0
4	Zombie	Live	Receive	0
5	Zombie	Live	Receive	0
6	Zombie	Live	Receive	0
7	Zombie	Live	Receive	0

III. Results

In this section, we apply the method described above to study two interesting population genetics problems: founders effect, and group competition and selection.

A. Founders Effect

Founder effect describes a certain type of population dynamics events where tribes grow from very small populations to large populations. The first paper describing the phenomenon was by Ernst Mayr in 1942 [7]. A much more recent paper was published by Ramachandran et al. [8].

We model the Founder effect in Mendel's Accountant by using a bottleneck event. The user may specify the starting point of the bottleneck, the growth rate of the population, the carrying capacity, and the population size threshold to trigger the fission event. We demonstrate eight tribes with a growth rate of 2x, starting population size of 2, population size fission threshold of 200. Here is a table of the results:

generation	pop size	tribes
1	2	1
2	4	1
3	8	1
4	16	1
5	32	1
6	64	1
7	128	1
8	256	2
9	512	4
10	1024	8
11	2048	8
12	4096	8
13	8192	8
14	16384	8

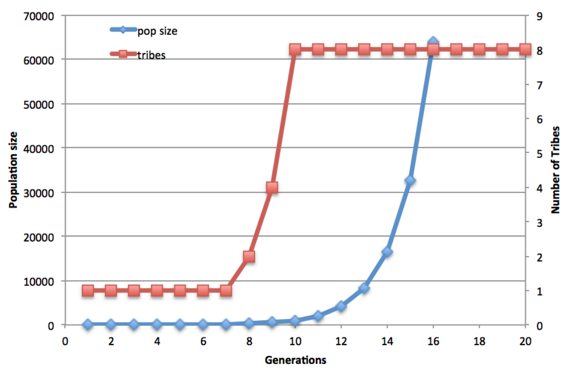

Figure 4. Population size and number of tribes for fission doubling event with 8 tribes and 2X growth rate.

This is shown in Fig. 4.

B. Radial Divergence Event

In this case, we model the tower of Babel event as shown in Fig. 5. We start with a dynamic growth from two individuals up to a carrying capacity of 1000 individuals, then we invoke a bottlenecke representing the flood down to six individuals. Then we invoke a radial fission event at generation 15, where the tribes split from one to four. Each tribe is assigned a different carrying capacity, and grows to its max carrying capacity.

IV. Conclusions

In this paper, we have discussed our work to develop a parallel tribal fission model using MPI. The model was used to demonstrate population genetics phenomenon such as founders effect and radial divergence. Mendel's Accountant is open source and can be freely downloaded at mendelsaccountant.info.

Acknowledgment

The authors would like to thank FMS Foundation for financial support for this study.

References

[1] Sanford, J., Baumgardner, J., Brewer, W., Gibson, P., & ReMine, W. "Mendel's Accountant: A biologically realistic forward-time population genetics program." *Scalable Computing: Practice and Experience* 8.2 (2001).

[2] Brewer, Wesley H., Franzine D. Smith, and J. Sanford. "Information loss: potential for accelerating natural genetic attenuation of RNA viruses." *Biological Information–New Perspectives. London: World Scientific* (2013): 369-84.

[3] Sanford, John, et al. "Using computer simulation to understand mutation accumulation dynamics and genetic load." *International Conference on Computational Science*. Springer Berlin Heidelberg, 2007.

[4] Sanford, John, Wesley Brewer, Franzine Smith, and John Baumgardner. "The waiting time problem in a model hominin population." *Theoretical Biology and Medical Modelling* 12, no. 1 (2015): 18.

[5] Saphir, William. "A Survey of MPI Implementations." *Lawrence Berkeley National Laboratory, University of California, Berkeley, CA, Nov. 6* (1997).

[6] Thompson, Elizabeth A. "Fission models of population variability." *Genetics* 93.2 (1979): 479-495.

[7] Mayr, Ernst. *Systematics and the origin of species, from the viewpoint of a zoologist*. Harvard University Press, 1942.

[8] Ramachandran, Sohini, et al. "Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa." *Proceedings of the National Academy of Sciences of the United States of America* 102.44 (2005): 15942-15947.

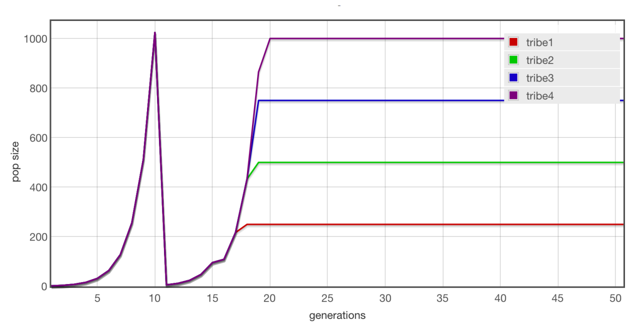

Figure 5. Radial fissioning scenario with non-homogenous tribes