4

Population Evolution in Hemophilia

Myung-Hoon Chung
Hongik University, Jochiwon
Korea

1. Introduction

Gene defects cause diseases. It is reasonable to argue that a disease is related to many genes and a gene affects on many diseases. This multiple-to-multiple mapping between genes and diseases makes it difficult to understand the roles of genes completely. However, in some case, only one gene is involved in one disease. A good example of one-to-one mapping is hemophilia related to gene F8 (National Hemophilia Foundation, 1998).

Biologists found that there are approximately 3 billion nitrogenous bases in 44+2 human DNA. Most sequences of these bases are irrelevant to genetic inheritance. Less than 3% of whole pair sequences are known to determine characteristic features of all human genes (Lander et al., 2001; Venter et al., 2001). Since the diversity of human's inherited characteristics is huge, it can only be explained by cooperation of multiple genes.

Single gene effect with multiple alleles such as blood groups (Chung et al., 1997), color blindness and hemophilia (Lee et al., 2001) can be studied by using the automata equations. Their solutions have some analogy with fixed points of renormalization group equations in physics (Mekjian, 1991; Perelson & Weisbuch, 1997), and lead into the Hardy-Weinberg formula (Haldane, 1935; Hedrick, 1985; Li, 1976).

We extended the automata equations to investigate multiple gene effects on population evolution with any number of loci and alleles in the presence of mutation and selection (Chung et al., 2003). As results of the study, we present the generalized Hardy-Weinberg formula and a simulation program on the Internet (Chung, 2007). The program explores simultaneous control of parameters that affect the behavior of gene variations in a population. We note that Hampe *et al.* (1998) studied population evolution in a different view point from ours. The advantage of our approach is that we do not need a large RAM memory because we do not treat individual person, but consider groups characterized by genes. One more good point beyond the work of Hampe *et al.* (1998) is that we achieve quickly equilibrium state. Using the simulation program, we find that the mortality rates due to gene inheritance are greatly enhanced for multiple gene cases. Another user friendly simulation program provides a convenient scheme for the most common cases (Quardokus, 2000). However, it is not easy to extend this scheme to include more sophisticated situations.

Human beings inherit not only biological genes but also social status from their parents. A few examples of the social inheritance are family name, nationality, or wealth. One of the outstanding questions is how or whether this social inheritance influences the standard biological gene evolution. We believe that this type of interaction may play a crucial

role in explaining appearance or disappearance of certain physical or social traits in some communities. In order to study population evolution with inheritance of social status, we treat inheritable social traits as *social genes* that behave similar to Mendelian genes. This approach allows us to handle the biological and the social genes in a unified way and to examine the mutual influence between the social and the biological genes.

One of the distinct features of the social gene is that it causes preferential non-random mating. Non-random mating exists even without the social gene concepts. Desire to avoid genetically inherited diseases and preferential sexual attraction may cause non-random mating. However, with the social restriction and prejudice, the tendency of non-random mating is expected to become more prominent and complicated.

The population evolution depends on four dominant factors imposed by nature with genes. These factors are (1) mating, (2) mutation, (3) reproduction, and (4) selection. The author proposed automata equations in order to describe the effects of these factors in the population evolution. These equations were used to determine the equilibrium population ratios in multiple gene inheritance, where arbitrary numbers of loci and alleles were allowed in the presence of natural selection, mutation and recombination. A user-friendly numerical simulation program was proposed for estimation of the infant mortality rate for fatal diseases (Chung et al., 2003). In order to incorporate the social genes in the inheritance scheme, we generalize the scheme to include non-random mating explicitly.

It is worthwhile to mention other population evolution studies, which are different from our approach. First of all, Cavalli-Sforza and Feldman (1981) developed a mathematically-rich theory of cultural transmission and evolution. Gene-culture coevolution refers to the evolutionary phenomena that arise from interactions between the biological and social inheritance systems (Aoki, 2001). In genetics, there exist current efforts to locate genes that contribute to diseases or to valuable traits (Piccolboni & Gusfield, 2003). Furthermore, it is essential to analyze the structure of populations on the basis of genetic data (Santafe et al., 2008). It is also well-known that the coalescent method (Hudson, 1991) is use to determine mutation rates (Thomson et al., 2000) and recombination rates (Fu & Li, 1999) in the way of statistical inference (Rosenberg & Nordborg, 2002). Studies on substructured populations introduce the similar feature of social gene, for instance, age (Charlesworth, 1994) or last name. We note that many studies related to mating already exist. The convergence of the multilocus systems under selection with a random mating was investigated (Nagylaki et al., 1999). In the population models for the diploid ancestral process with a random mating, the convergence criterium was proved (Möhle et al., 2003). Non-random mating has been found to play a significant role in the models of the population genetics. In the work related with a non-random mating (Hausken & Hirshleifer, 2001), the truthful signalling hypothesis was used in the mating competition theory. Strategic mating between males and females was also considered by Alpern and Reyniers (2005) and Radcliffe and Rass (1999).

The main purpose of this chapter is to provide a theoretical scheme and a simulation tool to handle the social and the biological genes in a unified way. We present a generalized numerical simulation tool to account for the role of the non-random mating induced by the social genes in addition to mutation, recombination and selection. We expect that the scheme will allow one to examine closely the impact of the social genes on the biological ones and vice versa. For example, we study hemophilia thoroughly without and with a social gene. Simulation results show that a medical screening to prevent birth of females with fatal

hemophilia increases the number of the male patients and the female carriers, thus clearly showing that the social effects significantly influence the inheritance of the biological gene.

2. Model

2.1 Symbols

The fundamental concept for the population evolution is the gene, which allows multiple loci, multiple allele inheritance with recombination and mutation. The l-th gene in the i-th chromosome is labeled by two indices (i,l). We denote $n^{(i,l)}$ as the number of alleles for the gene. Hence, we denote gene

$$G_a^{(i,l)},$$

where $1 \leq a \leq n^{(i,l)}$ and $i = 1,2,\cdots,22,X,Y,m,s$ including sex chromosome X,Y, mitochondria m, and social gene s. Here, we note that the social inheritance is treated on an equal footing with the biological one. However, the social genes may have different rules of reproduction and mutation rates, when compared with the biological genes.

The string of genes models each chromosome. We denote chromosome $C_a^{(i)}$ containing relevant L_i genes as

$$C_a^{(i)} = G_{a_1}^{(i,1)} G_{a_2}^{(i,2)} \cdots G_{a_{L_i}}^{(i,L_i)},$$

with $i = 1,2,\cdots,22,X,Y$. We write mitochondria M_a containing relevant L_m genes as

$$M_a = G_{a_1}^{(m,1)} G_{a_2}^{(m,2)} \cdots G_{a_{L_m}}^{(m,L_m)}.$$

Also we write social gene string S_a containing relevant L_s genes as

$$S_a = G_{a_1}^{(s,1)} : G_{a_2}^{(s,2)} : \cdots : G_{a_{L_s}}^{(s,L_s)}.$$

The number of distinguishable strings is given by $\prod_{l=1}^{L_i} n^{(i,l)} \equiv n^{(i)}$ with $i = 1,2,\cdots,22,X,Y,m,s$. It is useful to introduce chromosome pair

$$: C_{a_1}^{(i)}, C_{a_2}^{(i)} :$$

where the chromosome pair allele would run from 1 to $(n^{(i)}+1)n^{(i)}/2$ with $i = 1,2,\cdots,22,X$, while Y chromosome and mitochondria do not need the pairing.

Integrating all, a female genotype $T_a^{(F)}$ and a male genotype $T_a^{(M)}$ are expressed in terms of chromosome pairs, mitochondria and a series of social genes having multiple alleles. We denote for a female genotype,

$$T_a^{(F)} = (C_{p_1}^{(1)}, C_{q_1}^{(1)} : \cdots : C_{p_{22}}^{(22)}, C_{q_{22}}^{(22)} : C_{p_X}^{(X)}, C_{q_X}^{(X)} : M_{p_m} : S_{p_s}).$$

Similarly, we denote for a male genotype,

$$T_a^{(M)} = (C_{p_1}^{(1)}, C_{q_1}^{(1)} : \cdots : C_{p_{22}}^{(22)}, C_{q_{22}}^{(22)} : C_{p_X}^{(X)}, C_{q_Y}^{(Y)} : M_{p_m} : S_{p_s}).$$

We find that the number of different genotypes is given by $[\prod_{i=1}^{22} \frac{1}{2}(1+n^{(i)})n^{(i)}]\frac{1}{2}(1+n^{(X)})n^{(X)}n^{(m)}n^{(s)}$ for female, and $[\prod_{i=1}^{22} \frac{1}{2}(1+n^{(i)})n^{(i)}]n^{(X)}n^{(Y)}n^{(m)}n^{(s)}$ for male. Note that, for several given genes, the number of distinguishable genotypes increases dramatically.

For a genotype $T_a^{(S)}$ with $S = F$ or M, we introduce four kinds of population ratios at the n-th generation: adult, birth, parent before mutation, and effective parent after mutation, which are denoted by $A_n(T_a^{(S)})$, $B_n(T_a^{(S)})$, $P_n(T_a^{(F)} \times T_b^{(M)})$, and $\widetilde{P}_n(T_a^{(F)} \times T_b^{(M)})$, respectively. Here, the population ratios denote the frequencies of specific genotypes or genotype pairs of parents. We normalize so that the sums of any population ratios are equal to 1 as shown in Table 1. We shall notice the relationships between these population ratios later.

The population evolution is governed by the four factors: (1) mating, (2) mutation, (3) reproduction, and (4) selection. We introduce some symbols used in the population evolution as follows in each step.

In mating, $\omega(T_a^{(F)} \to T_b^{(M)})$ denotes the probability that a genotype $T_a^{(F)}$ adult woman mates with a $T_b^{(M)}$ adult man, while $\omega(T_b^{(M)} \to T_a^{(F)})$ is the probability that a genotype $T_b^{(M)}$ adult man mates with a $T_a^{(F)}$ adult woman. This mating probability reflects social and cultural effects. For the random mating case, the probability $\omega(T_a^{(F)} \to T_b^{(M)})$ is simply $A_n(T_b^{(M)})$, and $\omega(T_b^{(M)} \to T_a^{(F)}) = A_n(T_a^{(F)})$. Notice that $\omega(T_a^{(F)} \to T_b^{(M)}) \neq \omega(T_b^{(M)} \to T_a^{(F)})$.

In mutation, the genotype mutation rates $\mu(T_a^{(S)} \to T_b^{(S)})$ are written in terms of the gene mutation rates $\mu(G_c^{(i,l)} \to G_d^{(i,l)})$ and the frequency of recombination due to chromosomal crossover. It is known that the frequency of recombination between two locations depends on their distance. Mutation rates between chromosomes are given by

$$\mu(C_a^{(i)} \to C_b^{(i)}) = \prod_{l=1}^{L} \mu(G_{a_l}^{(i,l)} \to G_{b_l}^{(i,l)}).$$

The mutation rates of mitochondria $\mu(M_a \to M_b)$ and the mutation rates of social gene string $\mu(S_a \to S_b)$ are given as the above similarly. Since each chromosome behaves independently, we find the mutation rate of a specific genotype $T_a^{(S)}$ with the sex index $S = F$ or M as

$$\mu(T_a^{(S)} \to T_b^{(S)}) = [\prod_{i=1}^{22} \eta^{(i)}]\eta^{(S)}\mu(M_{a_m} \to M_{b_m})\mu(S_{a_s} \to S_{b_s}),$$

where $\eta^{(i)} = \eta^{(i)}(C_v^{(i)}, C_w^{(i)} \to C_p^{(i)}, C_q^{(i)})$ is given by

$$\eta^{(i)} = \begin{cases} \mu(C_v^{(i)} \to C_p^{(i)})\mu(C_w^{(i)} \to C_p^{(i)}) & \text{for } p = q \\ \mu(C_v^{(i)} \to C_p^{(i)})\mu(C_w^{(i)} \to C_q^{(i)}) + \mu(C_v^{(i)} \to C_q^{(i)})\mu(C_w^{(i)} \to C_p^{(i)}) & \text{for } p \neq q \end{cases}$$

Furthermore, $\eta^{(F)}(C_v^{(X)}, C_w^{(X)} \to C_p^{(X)}, C_q^{(X)})$ is the same as $\eta^{(i)}$, and

$$\eta^{(M)}(C_v^{(X)}, C_w^{(Y)} \to C_p^{(X)}, C_q^{(Y)}) = \mu(C_v^{(X)} \to C_p^{(X)})\mu(C_w^{(Y)} \to C_q^{(Y)}).$$

For the social genes, mutation rates are mainly determined by the society and will be different for each case. In consequence, all genotype mutation rates are written in terms of the gene mutation rates. After mutation, the effective parent population $\widetilde{P}_n(T_a^{(F)} \times T_b^{(M)})$ will make offsprings.

In reproduction, the reproduction coefficients $\xi(T_b^{(F)} \times T_c^{(M)} \to T_a^{(S)})$ are calculated with the assumption that randomly chosen half of the father's chromosomes and half of the mother's chromosomes are delivered to their baby, however only the mother's mitochondria becomes the baby's mitochondria. For the social genes, the rule of reproduction will be different on a case-by-case basis. Assuming equal preference for each genotype, the reproduction coefficients $\xi(T_b^{(F)} \times T_c^{(M)} \to T_a^{(S)})$ can be calculated. In fact, the reproduction coefficients are written as

$$\xi(T_a^{(F)} \times T_b^{(M)} \to T_c^{(S)}) = [\prod_{i=1}^{22} \zeta^{(i)}]\zeta^{(S)}\zeta^{(m)}\zeta^{(s)}.$$

Here, $\zeta^{(i)} = \zeta^{(i)}(C_v^{(i)}, C_w^{(i)} \times C_p^{(i)}, C_q^{(i)} \to C_s^{(i)}, C_t^{(i)})$ is determined by the following simple algorithm:

Start with $\zeta^{(i)} = 0$.

If $(s,t) = (\min(v,p), \max(v,p))$, then add $\frac{1}{4}$ to $\zeta^{(i)}$.

If $(s,t) = (\min(v,q), \max(v,q))$, then add $\frac{1}{4}$ to $\zeta^{(i)}$.

If $(s,t) = (\min(w,p), \max(w,p))$, then add $\frac{1}{4}$ to $\zeta^{(i)}$.

If $(s,t) = (\min(w,q), \max(w,q))$, then add $\frac{1}{4}$ to $\zeta^{(i)}$.

Hence, $\zeta^{(i)}$ is given by one of four values, 0, $\frac{1}{4}$, $\frac{1}{2}$, 1. Also, $\zeta^{(F)}(C_v^{(X)}, C_w^{(X)} \times C_p^{(X)}, C_q^{(Y)} \to C_s^{(X)}, C_t^{(X)})$ is similarly determined:

Start with $\zeta^{(F)} = 0$.

If $(s,t) = (\min(v,p), \max(v,p))$, then add $\frac{1}{2}$ to $\zeta^{(F)}$.

If $(s,t) = (\min(w,p), \max(w,p))$, then add $\frac{1}{2}$ to $\zeta^{(F)}$.

Furthermore, $\zeta^{(M)}(C_v^{(X)}, C_w^{(X)} \times C_p^{(X)}, C_q^{(Y)} \to C_s^{(X)}, C_t^{(Y)})$ is determined as follows:

Start with $\zeta^{(M)} = 0$.

If $(s,t) = (v,q)$, then add $\frac{1}{2}$ to $\zeta^{(M)}$.

If $(s,t) = (w,q)$, then add $\frac{1}{2}$ to $\zeta^{(M)}$.

For mitochondria, $\zeta^{(m)}(M_v \times M_p \to M_s)$ is determined as follows:

Start with $\zeta^{(m)} = 0$.

If $s = v$, then add 1 to $\zeta^{(m)}$.

Finally, the reproduction coefficients $\zeta^{(s)}$ for social genes will be determined by a case-by-case consideration.

Name	Symbol	Property
Adult population	$A_n(T_a^{(S)})$	$\sum_a A_n(T_a^{(S)}) = 1$
Birth population	$B_n(T_a^{(S)})$	$\sum_a B_n(T_a^{(S)}) = 1$
Parents population	$P_n(T_a^{(F)} \times T_b^{(M)})$	$\sum_{a,b} P_n(T_a^{(F)} \times T_b^{(M)}) = 1$
Effective parents population	$\widetilde{P}_n(T_a^{(F)} \times T_b^{(M)})$	$\sum_{a,b} \widetilde{P}_n(T_a^{(F)} \times T_b^{(M)}) = 1$
Mating probability	$\omega(T_a^{(F)} \to T_b^{(M)})$	$\sum_b \omega(T_a^{(F)} \to T_b^{(M)}) = 1$
Genotype mutation rate	$\mu(T_a^{(S)} \to T_b^{(S)})$	$\sum_b \mu(T_a^{(S)} \to T_b^{(S)}) = 1$
Reproduction coefficient	$\xi(T_a^{(F)} \times T_b^{(M)} \to T_c^{(S)})$	$\sum_c \xi(T_a^{(F)} \times T_b^{(M)} \to T_c^{(S)}) = 1$
Disadvantage factor	$\delta(T_a^{(S)})$	$0.0 \leq \delta(T_a^{(S)}) \leq 1.0$

Table 1. Symbols used in the paper. The index S in genotype represents female with $S = F$, and male with $S = M$.

In selection, since human beings with faulty genes have lower survival rate, we introduce disadvantage factor $\delta(T_a^{(S)})$ for each genotype, which is given by a value between 0 and 1. A larger value of disadvantage factor means less chance of survival. The value of 1 represents the complete extinction. The terminology of fitness can alternatively be used to replace the disadvantage factor.

2.2 Population equation

For given population ratios $A_n(T_a^{(S)})$ at the n-th generation, our prime concern is "what are the next generation population ratios?". To answer this question, we introduce the four main effects on population evolution: mating, mutation, reproduction, and selection. Based on these four events, using the parameters in relation to probability as explained in the previous subsection, we formulate automata equations for population evolution:

• Mating

$$P_n(T_a^{(F)} \times T_b^{(M)}) = A_n(T_a^{(F)})\omega(T_a^{(F)} \to T_b^{(M)}) = A_n(T_b^{(M)})\omega(T_b^{(M)} \to T_a^{(F)}), \tag{1}$$

• Mutation

$$\widetilde{P}_n(T_a^{(F)} \times T_b^{(M)}) = \sum_{c,d} P_n(T_c^{(F)} \times T_d^{(M)})\mu(T_c^{(F)} \to T_a^{(F)})\mu(T_d^{(M)} \to T_b^{(M)}), \tag{2}$$

• Reproduction

$$B_n(T_a^{(S)}) = \sum_{b,c} \widetilde{P}_n(T_b^{(F)} \times T_c^{(M)})\xi(T_b^{(F)} \times T_c^{(M)} \to T_a^{(S)}), \tag{3}$$

• Selection

$$A_{n+1}(T_a^{(S)}) = \frac{B_n(T_a^{(S)})\{1 - \delta(T_a^{(S)})\}}{1 - \sum_b B_n(T_b^{(S)})\delta(T_b^{(S)})}. \tag{4}$$

The first equation is similar to the detailed balance in Monte Carlo simulation. The second equation shows that new populations are written as the sum of mutations from old

populations. The birth populations are determined in the third equation. The fractional adult populations at the next generation after selection and normalization are given by the fourth equation. These automata equations enable us to calculate the evolution of the genotype frequencies.

For a given initial normalized set of adult populations, the automata equations will produce a fixed point of $A^*(T_a^{(S)})$ eventually. We note that the fixed point has a global stability, and is a function of mutation rates and disadvantage factors. We present the entire library for solving the population equation on the Internet (Chung, 2007). In fact, written in C# language, the single reusable library of Science.dll contains Science.Biology.PopulationGenetics, with which one can simulate all cases. It is open to the public, and runs on a personal computer with Windows operating system. Any number of loci and alleles, and any values of mutation rates and disadvantage factors are allowed simultaneously.

It is worth mentioning that these population equations are not universal. In fact, if we introduce age as a social gene, these equations should be modified because only certain aged adults can marry and reproduce offsprings.

2.3 Non-random mating

While it is difficult to obtain detailed information on the biological genes of a prospective marriage partner, it is easy to do so on the social genes. Thus, the mating involving social genes will be a non-random mating in general.

In the population equation for mating of Eq. (1) which is analogous to detailed balance in the Monte Carlo simulation, the sum of probabilities must be 1. In fact, we require

$$\sum_b \omega(T_a^{(F)} \to T_b^{(M)}) = \sum_a \omega(T_b^{(M)} \to T_a^{(F)}) = 1. \tag{5}$$

Introducing a so-called *mating factor* Γ_{ab} which should not be negative, we rewrite the parent population as

$$P_n(T_a^{(F)} \times T_b^{(M)}) = A_n(T_a^{(F)})\Gamma_{ab}A_n(T_b^{(M)}). \tag{6}$$

Here, Γ_{ab} are input parameters in this model. However, the condition of Eq. (5) assures that the elements of Γ_{ab} should satisfy

$$\sum_{a=1}^{n^{(F)}} A_n(T_a^{(F)})\Gamma_{ab} = 1 \text{ for } 1 \le b \le n^{(M)}, \tag{7}$$

$$\sum_{b=1}^{n^{(M)}} \Gamma_{ab}A_n(T_b^{(M)}) = 1 \text{ for } 1 \le a \le n^{(F)}, \tag{8}$$

$$\Gamma_{ab} \ge 0. \tag{9}$$

These restrictions on Γ_{ab} assume that all adults marry and reproduce offsprings. When some genotype adults have handicaps for marriage, it is reflected in the disadvantage factor in selection.

Since we obtain the same equation from Eqs. (7) and (8) as

$$\sum_{ab} A_n(T_a^{(F)})\Gamma_{ab}A_n(T_b^{(M)}) = 1 = \sum_a A_n(T_a^{(F)}) = \sum_b A_n(T_b^{(M)}), \tag{10}$$

the constraint equations of Eqs. (7) and (8) are not linearly independent. We note that the number of constraints is given by $n^{(M)} + n^{(F)} - 1$. Hence, the number of free input parameters is given by $n^{(F)}n^{(M)} - n^{(M)} - n^{(F)} + 1$.

Although there are many ways to assign free parameters, we present here four cases for the mating factors.

• Random mating

A trivial but important solution would be the random mating for which all of Γ_{ab} are given by 1.

• Selective mating

If only one specific genotype is preferred completely, say $T_1^{(F)}$ and $T_1^{(M)}$, we let the free parameters of mating factors all zero as

$$(\Gamma_{ab}) = \begin{pmatrix} \Gamma_{11} & \Gamma_{12} & \Gamma_{13} & \cdots & \Gamma_{1n^{(M)}} \\ \Gamma_{21} & 0 & 0 & & 0 \\ \Gamma_{31} & 0 & 0 & & 0 \\ \vdots & & & \ddots & \vdots \\ \Gamma_{n^{(F)}1} & 0 & 0 & \cdots & 0 \end{pmatrix}. \tag{11}$$

In order to satisfy the constraints, other $n^{(M)} + n^{(F)} - 1$ factors should be given by

$$(\Gamma_{ab}) = \begin{pmatrix} \frac{-1+A_n(T_1^{(F)})+A_n(T_1^{(M)})}{A_n(T_1^{(F)})A_n(T_1^{(M)})} & \frac{1}{A_n(T_1^{(F)})} & \frac{1}{A_n(T_1^{(F)})} & \cdots & \frac{1}{A_n(T_1^{(F)})} \\ \frac{1}{A_n(T_1^{(M)})} & 0 & 0 & & 0 \\ \frac{1}{A_n(T_1^{(M)})} & 0 & 0 & & 0 \\ \vdots & & & \ddots & \vdots \\ \frac{1}{A_n(T_1^{(M)})} & 0 & 0 & \cdots & 0 \end{pmatrix}. \tag{12}$$

We note that the mating factors depend on population ratios at each generation. It should be emphasized from Γ_{11} that the population of $A_n(T_1^{(F)}) + A_n(T_1^{(M)})$ must be greater than 1 at the initial stage, and remains greater than 1 until the population ratios arrive at the equilibrium in this non-random mating.

• Hierarchical mating

For the case of $n^{(F)} = n^{(M)}$, we can define the mating factors as

$$(\Gamma_{ab}) = \begin{pmatrix} \Gamma_{11} & 0 & 0 & \cdots & 0 \\ \Gamma_{21} & \Gamma_{22} & 0 & & 0 \\ 0 & \Gamma_{32} & \Gamma_{33} & & 0 \\ \vdots & & & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \Gamma_{n^{(F)}n^{(M)}} \end{pmatrix}. \tag{13}$$

Here all the free input parameters of the mating factors are given by zero. The $n^{(M)} + n^{(F)} - 1$ factors should be determined by the constraints. As a result, we can determine Γ_{ab} as

$$\Gamma_{11} = \frac{1}{A_n(T_1^{(M)})}, \tag{14}$$

$$\Gamma_{21} = \frac{1}{A_n(T_2^{(F)})}(1 - \frac{A_n(T_1^{(F)})}{A_n(T_1^{(M)})}), \tag{15}$$
$$\cdots \quad ,$$

where we omit presenting other Γ_{ab}, which have similar expressions in terms of the genotype frequencies.

• General non-random mating

We consider a new and more general approach by introducing the method of the linear programming.

When non-random mating is involved, there are $n^{(F)} \cdot n^{(M)}$ unknown values of Γ_{ab}. Since the number of the unknowns is bigger than the number of the restriction equations of Eqs. (7) and (8), there is no unique way to determine the mating factors Γ_{ab}. However, if we add a theoretical restriction such as maximization or minimization of a specific population, a unique process becomes possible. We introduce an additional condition of maximizing Z, which is written in terms of the parent populations as

$$Z = \sum_{ab} \beta_{ab} P_n(T_a^{(F)} \times T_b^{(M)}), \tag{16}$$

where the parameters β_{ab} represent the tendency of mating. In the process of maximizing Z, a given positive (negative) value of β_{ab} will produce a bigger (smaller) value of $P_n(T_a^{(F)} \times T_b^{(M)})$, which corresponds to inbreeding (outbreeding) between $T_a^{(F)}$ type female and $T_b^{(M)}$ type male. This problem of maximizing the objective function Z with the constraints of Eqs. (7)-(9) is well known, and can be solved as a linear programming problem, for example by the simplex method (Press et al., 1992).

The nature of β_{ab} values can be understood if we use practical demographic data, for example, incomes or divorce rates for parent groups. For instance, for the society that wants to reduce the divorce rates, it is possible to find the mating factors Γ_{ab} for the minimal divorce rate. If β_{ab} values represent incomes, we find the mating factors for the maximum income of the society. In this paper, we consider only medical screening.

3. Hemophilia

For hemophilia, the locus of the relevant gene called F8 is on X chromosome, and the gene has two alleles: normal X and abnormal X'. For female, three genotypes exist: normal XX, carrier XX', and disease $X'X'$. For male, there are two genotypes: normal XY, and disease $X'Y$. In the model of hemophilia, it is worthwhile to present all input parameters explicitly. First of

all, for the random mating, all of mating factors are given by 1 as

$$\Gamma_{ab} = 1.$$

The mutation rates are given by

$$\mu(X \to X') = \alpha,\ \ \mu(X \to X) = 1 - \alpha,$$
$$\mu(X' \to X) = \beta,\ \ \mu(X' \to X') = 1 - \beta.$$

The reproduction coefficients are found as

$$\xi(XX \times XY \to XX) = 1,\ \ \xi(XX \times XY \to XX') = 0,\ \ \xi(XX \times XY \to X'X') = 0,$$
$$\xi(XX \times X'Y \to XX) = 0,\ \ \xi(XX \times X'Y \to XX') = 1,\ \ \xi(XX \times X'Y \to X'X') = 0,$$
$$\xi(XX' \times XY \to XX) = 0.5,\ \ \xi(XX' \times XY \to XX') = 0.5,\ \ \xi(XX' \times XY \to X'X') = 0,$$
$$\xi(XX' \times X'Y \to XX) = 0,\ \ \xi(XX' \times X'Y \to XX') = 0.5,\ \ \xi(XX' \times X'Y \to X'X') = 0.5,$$
$$\xi(X'X' \times XY \to XX) = 0,\ \ \xi(X'X' \times XY \to XX') = 1,\ \ \xi(X'X' \times XY \to X'X') = 0,$$
$$\xi(X'X' \times X'Y \to XX) = 0,\ \ \xi(X'X' \times X'Y \to XX') = 0,\ \ \xi(X'X' \times X'Y \to X'X') = 1,$$
$$\xi(XX \times XY \to XY) = 1,\ \ \xi(XX \times XY \to X'Y) = 0,$$
$$\xi(XX \times X'Y \to XY) = 1,\ \ \xi(XX \times X'Y \to X'Y) = 0,$$
$$\xi(XX' \times XY \to XY) = 0.5,\ \ \xi(XX' \times XY \to X'Y) = 0.5,$$
$$\xi(XX' \times X'Y \to XY) = 0.5,\ \ \xi(XX' \times X'Y \to X'Y) = 0.5,$$
$$\xi(X'X' \times XY \to XY) = 0,\ \ \xi(X'X' \times XY \to X'Y) = 1,$$
$$\xi(X'X' \times X'Y \to XY) = 0,\ \ \xi(X'X' \times X'Y \to X'Y) = 1.$$

In the United States, about 17,000 people have hemophilia. About one in 7,500 live male births has hemophilia and about one in 25,000,000 live female births has hemophilia. From the data, we find that almost all female babies with hemophilia are dead during the pregnancy. Thus, it is natural to assign the disadvantage factors as

$$\delta(XX) = \delta(XX') = \delta(XY) = 0,\ \delta(X'X') = 1,\ \delta(X'Y) = \delta.$$

Note that we take $\delta = 0$ if male patients are completely cured. In the next subsections, we study the effect of male patient treatment by changing δ. The value of δ reflects the social development and standards of the general health care of particular geographical region where the hemophilia population is living.

3.1 Random mating with $\delta = 1$

The most natural case of hemophilia would be random mating with fatal mortality. In this case, we let $\delta(X'Y) = \delta = 1$. We now focus on mutation rates $\mu(X \to X') = \alpha$ and $\mu(X' \to X) = \beta$.

In fact, we determine the mutation rate $\mu(X \to X')$ as about 3.3×10^{-5} using the demographic data of birth population, $XY : X'Y = 0.9999 : 0.0001$. The population equation at $\alpha = 3.3 \times 10^{-5}$ shows that the equilibrium adult population is calculated as

$$\begin{aligned} XX : XX' : X'X' &= 0.999868 : 0.000132 : 0, \\ XY : X'Y &= 1 : 0, \end{aligned} \tag{17}$$

where the value of 0.000132 was rather insensitive on the value of β. We found that this low sensitivity originates from the severe disadvantage factor in hemophilia. Although there is a report for existence of a gene self-repairing mechanism in the process of evolution (Avise, 1993), we assume that self-repairing is very rare. Throughout this section hereafter, we simply let $\beta = 0.0$ and $\alpha = 3.3 \times 10^{-5}$ for hemophilia.
It is shown that these equilibrium population ratio values can be easily achieved not only analytically but also numerically (Lee et al., 2001). Numerical simulations show that these values become stable after about 100 generations, having no dependence on the initial state.

3.2 Random mating with $\delta = 0.1$

We consider the case where male patients of hemophilia can be treated so that some of them could lead normal lives, although a genetic defect remains intact. However, since female baby patients are dead during the pregnancy, it is assumed that there is still no cure for female patients as it stands at present. In fact, when the disadvantage factors are given by $\delta(XX) = \delta(XX') = \delta(XY) = 0$, $\delta(X'Y) = 0.1$, and $\delta(X'X') = 1$ with the same mutation rates as the above, we find the equilibrium population:

$$\begin{aligned} XX : XX' : X'X' &= 0.998121 : 0.001879 : 0, \\ XY : X'Y &= 0.999125 : 0.000875, \end{aligned} \tag{18}$$

in the random mating case where all $\Gamma_{ab} = 1$. These numbers will be used as references to study the effect of the non-random mating below.
Recently an interesting article (Stonebraker et al., 2010) on hemophilia prevalence in different countries was published. The prevalence (per 100,000 males) for high income countries was 12.8 ± 6.0 (mean $\pm$ SD) whereas it was 6.6 ± 4.8 for the rest of the world. Within a country, there was a strong trend of increasing prevalence over time: the prevalence for Canada ranged from 10.2 in 1989 to 14.2 in 2008 and for the United Kingdom it ranged from 9.3 in 1974 to 21.6 in 2006. The data are consistent with the fact that the cure results in increasing hemophilia prevalence.

3.3 Non-random mating with $\delta = 0.1$ in minimizing $\mathbf{P}_n(\mathbf{XX'} \times \mathbf{X'Y})$

Since rapid advances in molecular genetics have highlighted the potential use of genetic testing to screen for adult-onset chronic diseases (Burke et al., 2001), medical screening becomes more plausible.
We consider an artificial circumstance, where society-wide medical screening is available to control the mating factors. For example, we consider the case of hemophilia with a medical screening. The goal in this consideration is to find the equilibrium populations in that society. We assume that the medical screening helps male patients avoid marrying disease carrier

women. This restriction lead us to choose the objective function as

$$Z = -P_n(XX' \times X'Y),$$

with which we expect a minimal value of $P_n(XX' \times X'Y)$ for maximizing Z. In consequence, with the previous disadvantage factors and mutation rates, the population equations give the equilibrium as

$$\begin{aligned} XX : XX' : X'X' &= 0.998089 : 0.001911 : 0, \\ XY : X'Y &= 0.999110 : 0.000890. \end{aligned} \tag{19}$$

The corresponding objective function Z is given by simply 0 as expected. We find the different numbers of the male patients and the female carriers from those of the random mating in Eq. (18).

3.4 Non-random mating with $\delta = 0.1$ in maximizing $\mathbf{P}_n(\mathbf{XX'} \times \mathbf{X'Y})$

It is instructive to compare the above result with that of the opposite situation where the objective function is given by

$$Z = P_n(XX' \times X'Y)$$

to maximize $P_n(XX' \times X'Y)$. For the same disadvantage factors and mutation rates as the above, the population equation gives the equilibrium as

$$\begin{aligned} XX : XX' : X'X' &= 0.999868 : 0.000132 : 0, \\ XY : X'Y &= 0.999911 : 0.000089. \end{aligned} \tag{20}$$

The corresponding objective function Z at equilibrium is given by 8.90937×10^{-5}, which means that all male patients marry with female carriers because $X'Y$ is given by 0.000089 as shown in the above.

We find that the population ratios of Eq. (18) in random mating are between two extremal cases of Eqs. (19) and (20), which correspond to outbreeding and inbreeding of female carriers and male patients, respectively.

3.5 Non-random mating in association of a social gene

We now consider the effect of the social inheritance on the evolution of the biological genes in the case of hemophilia. We consider a two-gene system, where one of the relevant genes is hemophilia on X chromosome, and the other is a social gene. For simplicity, we assume that the social gene has two allele: Rich(R) and Poor(P). In this model, there are six female genotypes and four male genotypes:

$$\begin{aligned} T_1^{(F)} &= XX : R, \quad T_2^{(F)} = XX : P, \\ T_3^{(F)} &= X'X : R, \quad T_4^{(F)} = X'X : P, \\ T_5^{(F)} &= X'X' : R, \quad T_6^{(F)} = X'X' : P, \\ T_1^{(M)} &= XY : R, \quad T_2^{(M)} = XY : P, \\ T_3^{(M)} &= X'Y : R, \quad T_4^{(M)} = X'Y : P. \end{aligned} \tag{21}$$

For the biological gene, we use the same mutation rates $\mu(X \to X')$, $\mu(X' \to X)$ and the same reproduction rates $\xi^{(b)}$ as before.
For the social gene, we let the mutation rates be

$$\mu(R \to P) = 0.2 \text{ and } \mu(P \to R) = 0.1 \tag{22}$$

as a sample example. Also, we assign the reproduction rates $\xi^{(s)}$ with which the whole reproduction rates ξ in the population equations are written as $\xi = \xi^{(b)}\xi^{(s)}$:

$$\begin{aligned}
&\xi^{(s)}(R \times R \to R) = 1, \ \xi^{(s)}(R \times R \to P) = 0,\\
&\xi^{(s)}(R \times P \to R) = 1, \ \xi^{(s)}(R \times P \to P) = 0,\\
&\xi^{(s)}(P \times R \to R) = 1, \ \xi^{(s)}(P \times R \to P) = 0,\\
&\xi^{(s)}(P \times P \to R) = 0, \ \xi^{(s)}(P \times P \to P) = 1.
\end{aligned} \tag{23}$$

We further assign the disadvantage factors as

$$\begin{aligned}
\delta(T_1^{(F)}) &= 0, \quad \delta(T_2^{(F)}) = 0,\\
\delta(T_3^{(F)}) &= 0, \quad \delta(T_4^{(F)}) = 0,\\
\delta(T_5^{(F)}) &= 1, \quad \delta(T_6^{(F)}) = 1,\\
\delta(T_1^{(M)}) &= 0, \quad \delta(T_2^{(M)}) = 0,\\
\delta(T_3^{(M)}) &= 0, \quad \delta(T_4^{(M)}) = 0.1.
\end{aligned} \tag{24}$$

These factors imply that hemophilia is still fatal for female patients, while the rich social gene can make male patients lead normal lives.
In order to analyze equilibrium population resulted from inbreeding between the normal and rich genotypes, we set the objective function as

$$Z = P_n(T_1^{(F)} \times T_1^{(M)}).$$

With all the given parameters as the above with the inbreeding, we find the equilibrium population ratios as

$$\begin{aligned}
&T_1^{(F)} : T_2^{(F)} : T_3^{(F)} : T_4^{(F)} : T_5^{(F)} : T_6^{(F)}\\
&\quad = 0.833195 : 0.154462 : 0.009700 : 0.002643 : 0 : 0,\\
&T_1^{(M)} : T_2^{(M)} : T_3^{(M)} : T_4^{(M)}\\
&\quad = 0.837755 : 0.156138 : 0.005223 : 0.000884.
\end{aligned} \tag{25}$$

We clearly observe that the existence of social gene drastically influences the inheritance of the biological genes. We believe that this influence is induced by the disadvantage factor of the genotype, $T_4^{(M)}$.
In order to examine the role of the disadvantage factor in the evolution mechanism, we have carried out a calculation by changing only $\delta(T_4^{(M)})$ from 0.1 to 1. The resulting population

ratios are given by

$$
\begin{aligned}
&T_1^{(F)} : T_2^{(F)} : T_3^{(F)} : T_4^{(F)} : T_5^{(F)} : T_6^{(F)} \\
&= 0.826752 : 0.171908 : 0.001127 : 0.000213 : 0 : 0, \\
&T_1^{(M)} : T_2^{(M)} : T_3^{(M)} : T_4^{(M)} \\
&= 0.827356 : 0.172040 : 0.000604 : 0.
\end{aligned} \tag{26}
$$

Comparing with Eq. (25), we clearly observe that the change of the disadvantage factor has a significant influence on the hemophiliac biological evolution. We also note that the ratio of rich to poor is higher in the diseased than in the healthy, clearly indicating the social gene effect of richness. Thus, we note again that existence of a particular social gene can have significant effect on the evolution process.

We note that a drawback of our approach is in the large number of tunable parameters. While mutation rates and recombination rates for biological genes may be adjusted by the coalescent method, it is hard to find an empirical basis for social genes. The social gene of last name probably gains some empirical support. However, if no support is provided, it is reasonable to perform sensitivity analysis by scanning around chosen parameters.

There are many genetic diseases other than hemophilia in the real world. As an example, the same approach may be applied to the case of congenital hypothyroidism (Calaciura et al., 2002). Expansion of the present scheme to other cases including more diverse social genes remains as a future study.

4. Conclusion

We have presented a theoretical scheme and a simulation program to study population genetics in a realistic complex situation. The result shows that the multiplicity of the gene loci greatly affect the demographic distribution of fractional population ratios. We suggest that more detailed demographic data including gene mutation (Porter, 1968; Strachan, 1996) and fitness is desirable to elaborate the theory further.

Treating social status as an inheritance trait, we introduce the concept of the social gene. Treating the social genes on equal footing with the biological genes, we build a unified framework to find the population evolution (Cavalli-Sforza & Bodmer, 1996) in the biological and the social inheritance. The nature of the social inheritance inevitably introduces the concept of non-random mating. This framework is used to investigate the detailed cases of hemophilia.

For hemophilia, we have presented a theoretical scheme to determine the mating factors uniquely in the non-random mating process by introducing the objective function concept. Finally, we have considered the effect of the social gene with two alleles, Rich and Poor on the biological gene of hemophilia. The result shows that the introduction of a social gene or social inheritances change the population evolution of the biological gene. We note that the social gene concept introduced here is different from polyphenism, for which multiple phenotypes can arise from a single genotype as a result of differing environment conditions. The reason for the difference is that the social gene is inherited with the biological gene, whereas the environment is not.

The current study does not include any dynamic change of mutation factors, which are expected especially in the social gene inheritance. The collective and the dynamic behavior are subjects of future studies.

5. References

Alpern, S. & Reyniers, D. (2005). Strategic mating with common preferences. *Journal of Theoretical Biology*, 237 : 337-354.

Aoki, K. (2001). Theoretical and empirical aspects of gene-culture coevolution. *Theoretical Population Biology*, 59 : 253-261.

Avise, J.C. (1993). The evolutionary biology of aging, sexual reproduction, and DNA repair. *Evolution*, 47 : 1293-1301.

Bamshad, M.J. et al. (1998). Female gene flow stratifies hindu castes. *Nature*, 395 : 651-652.

Burke, W.; Coughlin, S.S.; Lee, N.C.; Weed, D.L. & Khoury, M.J. (2001). Application of population screening principles to genetic screening for adult-onset conditions. *Genetic Testing*, 5 : 201-211.

Calaciura, F. et al. (2002). Genetics of specific phenotypes of congenital hypothyroidism: a population-based approach. *Thyroid*, 12 : 945-951.

Cavalli-Sforza, L.L. & Bodmer, W.F. (1971). *The Genetics of Human Populations*, W. H. Freeman and Company, San Francisco.

Cavalli-Sforza, L.L. & Feldman, M.W. (1981). C*ultural Transmission and Evolution: A quantitative approach*, Princeton University Press, New Jersey.

Charlesworth, B. (1994). *Evolution in age-structured populations second edition*, Cambridge University Press, Cambridge.

Chung, M.-H.; Lee, S.P.; Kim, C.K. & Nahm, K. (1997). Fractional populations of blood groups. *Physical Review E*, 56 : 865-869.

Chung, M.H.; Kim, C.K. & Nahm, K. (2003). Fractional populations in multiple gene inheritance. *Bioinformatics*, 19 : 256-260.

Chung, M.H. (2007). http://www.sciencecode.com/populationgenetics.htm.

Chung, M.H.; Kim, C.K. (2010). Non-random mating involving inheritance of social status. *Journal of Computational Biology*, 17 : 745-754.

Fu, Y.X. & Li, W.H. (1999). Minireview: coalescing into the 21st century: an overview and prospects of coalescent theory. *Theoretical Population Biology*, 56 : 1-10.

Haldane, J.B.S. (1935). The rate of spontaneous mutation of a human gene. *Journal of Genetics*, 31 : 317-326.

Hampe, J.; Wienker, T.; Schreiber, S. & Nürnberg, P. (1998). POPSIM: a general population simulation program. *Bioinformatics*, 14 : 458-464.

Hausken, K. & Hirshleifer, J. (2001). The truthful signalling hypothesis: an explicit general equilibrium model. *Journal of Theoretical Biology*, 228 : 497-511.

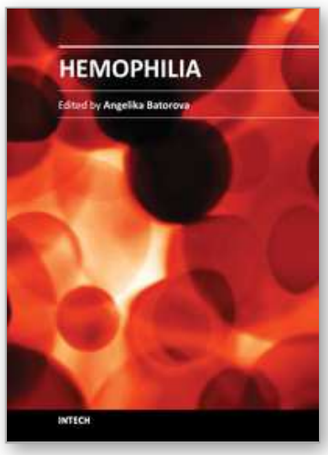
Hedrick, P.W. (1985). *Genetics of Populations*, Jones And Bartlett Publishers, Boston.

Hudson, R.R. (1991). *Oxford Surveys in Evolutionary Biology Vol. 7*, Cambridge University Press, Cambridge, pp. 1-44.

Lander, E.S. et al. (2001). Initial sequencing and analysis of the human genome. *Nature*, 409 : 860-921.

Lee, S.P.; Chung, M.-H.; Kim, C.K. & Nahm, K. (2001). Fractional populations in sex-linked inheritance. *Physica A*, 291 : 533-541.

Li, C.C. (1976). *First Course in Population Genetics*, Boxwood, Pacific Grove.
Mekjian, D.A.Z. (1991). Cluster distributions in physics and genetic diversity. *Physical Review A*, 44 : 8361-8374.
Möhle, M. & Sagitov, S. (2003). Coalescent patterns in exchangeable diploid population models. *Journal of Mathematical Biology*, 47 : 337-352.
Nagylaki, T.; Hofbauer, J. & Brunovský, P. (1999). Convergence of multilocus systems under weak epistasis or weak selection. *Journal of Mathematical Biology*, 38 : 103-133.
National Hemophilia Foundation. (1998). http://www.hemophilia.org.
Perelson, A.S. & Weisbuch, G. (1997). Immunology for physicists. *Reviews of Modern Physics*, 69 : 1219-1267.
Piccolboni, A. & Gusfield, D. (2003). On the Complexity of Fundamental Computational Problems in Pedigree Analysis. *Journal of Computatinal Biology*, 10 : 763-774.
Porter, I.H. (1968). *Heredity and Disease*, McGraw-Hill, New York.
Press, W.H.; Teukolsky, S.A.; Vetterling, W.T. & Flannery, B.P. (1992). *Numerical recipes in C second edition*, Cambridge University Press, Cambridge, pp. 430-444.
Quardokus, E. (2000). Modeling population genetics. *Science*, 288 : 458-459.
Radcliffe, J. & Rass, L. (1999). Strategic and genetic models of evolution. *Mathematical Biosciences*, 156 : 291-307.
Rosenberg, N.A. & Nordborg, M.N. (2002). Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. *Nature Reviews Genetics*, 3 : 380-390.
Santafe, G.; Lozano, J.A. & Larranaga, P. (2008). Inference of Population Structure Using Genetic Markers and a Bayesian Model Averaging Approach for Clustering. *Journal of Computational Biology*, 15 : 207-220.
Stonebraker, J.S.; Bolton-Maggs, P.H.B.; Michael Soucie, J.; Walker, I. & Brooker, M. (2010). A study of variations in the reported haemophilia A prevalence around the world. *Haemophilia*, 16 : 20-32.
Strachan, T. & Read, A.P. (1996). *Human Molecular Genetics*, Bios Scientific Publishers, Oxford.
Thomson, R.; Pritchard, J.K.; Shen, P.; Oefner, P.J. & Feldman, M.W. (2000). *Proceedings of the National Academy of Sciences*, 97 : 7360-7365.
Venter, J.C. et al. (2001). The sequence of the human genome. *Science*, 291 : 1304-1351.



Hemophilia

Edited by Dr. Angelika Batorova

ISBN 978-953-51-0429-2
Hard cover, 130 pages
Publisher InTech
Published online 30, March, 2012
Published in print edition March, 2012

This book demonstrates the great efforts aimed at further improving the care of the hemophilia, which may bring further improvement in the quality of life of hemophilia persons and their families.

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

IntechOpen