Research article

The origin of complex human diversity: Stochastic epistatic modules and the intrinsic compatibility between distributional robustness and phenotypic changeability

Shinji Ijichi[1,2,*], Naomi Ijichi[2], Yukina Ijichi[2], Chikako Imamura[3], Hisami Sameshima[1], Yoichi Kawaike[1], Hirofumi Morioka[1]

[1] Health Service Center, Kagoshima University, 1-21-24 Korimoto, Kagoshima 890-8580, Japan
[2] Institute for Externalization of Gifts and Talents, 7421-1 Shimofukumoto, Kagoshima 891-0144, Japan
[3] Support Center for Students with Disabilities, Kagoshima University, 1-21-30 Korimoto, Kagoshima 890-0065, Japan

*Correspondence: jiminy@hsc.kagoshima-u.ac.jp (Shinji Ijichi)

Abstract

The continuing prevalence of a highly heritable and hypo-reproductive extreme tail of a human neurobehavioral quantitative diversity suggests the reproductive majority retains the genetic mechanisms for extremes. From the perspective of stochastic epistasis, the effect of an epistatic modifier variant can randomly vary in both phenotypic value and effect direction among carriers depending on the genetic identity and the modifier carriers are ubiquitous in the population. The neutrality of the mean genetic effect in carriers ensures the survival of the variant under selection pressures. Functionally or metabolically related modifier variants make an epistatic network module and dozens of modules may be involved in the phenotype. To assess the significance of stochastic epistasis, a simplified module-based model was simulated. The individual repertoire of the modifier variants in a module also contributes in genetic identity, which determines the genetic contribution of each carrier modifier. Because the entire contribution of a module to phenotypic outcome is unpredictable in the model, the module effect represents the total contribution of related modifiers as a stochastic unit in simulations. As a result, the intrinsic compatibility between distributional robustness and quantitative changeability could mathematically be simulated using the model. The artificial normal distribution shape in large-sized simulations was preserved in each generation even if the lowest fitness tail was non-reproductive. The robustness of normality across generations is analogous to the real situation of complex human diversity, including neurodevelopmental conditions. The repeated regeneration of a non-reproductive extreme tail may be essential for survival and change of the reproductive majority, implying extremes for others. Further simulation to illustrate how the fitness of extreme individuals can be low across generations may be necessary to increase the plausibility of this stochastic epistasis model.

Keywords

Neurodevelopmental conditions; autism; schizophrenia; stochastic variation; complex trait; population diversity; quantitative change

Submitted: January 23, 2017; Accepted: April 5, 2017

1. Introduction

In quantitative complex traits, offspring trait value is determined by genes inherited from parents (nature), environmental factors (nurture), and the mixture (gene-environment interactions) [1]. A balanced perspective, that appreciates both nature and nurture, is important and valid for the developmental trajectories of each phenotype, as well as for the eventual outcomes [2]. Although the presence of unpredictability or stochasticity is easily predictable in phenotypic variations associated with complex environmental factors and gene-environment interaction [3], non-environmentally heritable stochasticity has attracted little investigative attention. Traditional intrinsic stochastic phenomena including the usual gene expression noise and epigenetically-driven monoallelic expression cannot be included in heritable stochasticity, because the phenotypic outcomes or patterns in genetically identical pairs are discordant and usually not replicable in offspring [4, 5]. In principle, the confirmation of

heritable unpredictability or stochasticity is only available from the experimental conditioning of genetic backgrounds [6–8]. Therefore, high concordance in monozygotic twins with a disparity between monozygotic and dizygotic concordance rates may be the only clue to predict the presence of heritable unpredictability. Non-environmental heritable stochasticity was first employed to explain the survival of a hypo-reproductive extreme tail (autism) of a complex human behavioral diversity [9]. The survival of this highly heritable and mainly sporadic behavioral condition [10] dictates an epistasis-associated stochastic fitness oscillation across generations (phenotypic trade-offs) and unpredictability of genetic effects [9]. Importantly, the supposed genetic underpinning can load the extreme tail with disabilities and allow the residual majority to enjoy their normal functions, and the genetic factors are not necessary to explain the presence of extreme tails for the stochastic epistasis concept [9]. In the inevitably stochastic process, whose phenotypic outcomes are

concordant in genetically identical pairs, the phenotypic value of a related modifier variant randomly changes with each generation and varies in a generation because of genetic identity [9, 11–13]. As a result, the mean fitness values of each modifier variant and the mean total fitness of an epistatic network (module), whose constituents are the modifiers and one or more of the evolutionarily survived (monomorphic) switches for stochasticity, can be net neutral both across generations and within a single generation [9, 11, 13]. Phenotypic stochasticity can render modifiers and modules hidden or missing, while hidden genetic architectures can maintain phenotypic diversity under selection [13]. Modest correlations between core characteristics in a complex trait and the sizable genetic overlap among multiple distinct complex conditions can be illustrated by the infidelity of pleiotropy in stochastic epistasis [9, 12, 13]. This theoretical challenge originates in an insight into the robustness of fitness-associated phenotypic diversity across generations, with the core of the concept being the presence of fitness neutrality resulting from the stochasticity of genetic effect [9]. Modifier variants that alter the noise of gene expression can be detected by comparative studies of experimental cell strains [14–16], and an evolvable stochasticity in epigenetic variation has been revealed using tissue samples [17]. Without reciprocal citation of each other's work, incomplete penetrance of complex human conditions and unpredictability of disease susceptbility has been deductively reasoned using cell-level stochasticity [15, 16], while evolvable epigenetic noise has been considered as the driving force of development, differentiation, evolutionary adaptation, and disease susceptibility [17].

The stochastic epistasis model never dismisses the comprehensive view that appreciates the involvement of the genetic effects of major variants, environmental factors, and gene-environment interactions [9, 11, 13]. Cis-acting genetic interactions associated with genetic imprinting is also considered as one of the related inter-gene interactions in the stochastic epistatic modules [12]. Because the need for an enormous assortment of physiological responses in complexity and specificity depends upon stochasticity-based phenotypic diversity [18], stochastic epistasis may be critical in complex human traits that include neurodevelopmental quantitative conditions [9, 11]. These complex traits involve the degree of phenotypic penetrance of a major variant gene effect, the liability to dichotomous diagnosis, the latent period length of late-onset diseases, clinical response to treatment, and resistance/vulnerability to pathogens, toxins, or mental stress. Stochastic epistasis may have an especially important role in complex conditions which have a disparity between monozygotic and dizygotic concordances in twin studies like autism and schizophrenia. In this article, a simplified module-based model was employed to characterize stochastic epistasis.

2. Methods

The stochastic effects of the related epistatic modifier variants on complex traits and the correlative genetic contribution of the parents were simulated with a simplified model. Because the related modifiers stochastically affect the phenotype according to the genetic identity as a network (module) and the effect values of related modules are also stochastic [11], the summation of stochastic module effects was added as the stochastic epistatic component according to the *de novo* genetic identity in the zygote including germinal mutations and chromosomal recombination and shuffling. Genetic effects of major variants and the environmental contribution were not included

in this model so as to exclusively evaluate the nature of the stochastic epistasis. In preliminary models, the number of related epistatic modules was documented by the number of evolutionarily surviving monomorphic switches for the stochastic epistasis and modifier variants were included in the random genetic individualities [13]. It has previously been demonstrated that the normal distribution of a quantitative trait can be well demonstrated by this model [11, 13]. The stochasticity of the component is supposed to be underpinned by sufficiently randomized genetic identity. The phenotypic value of an offspring (Xo) was obtained from the following formula.

$$Xo = a \times Xp + b \times Xm + c \times \sum_{i=1}^{m} Ui$$

The contribution from each stochastic module (Ui) ranges from -1 to 1 (actually 0.999999999) using the uniform random number generator of a spread sheet application. As described above, a stochastic module can generate an unpredictable genetic effect and previous preliminary simulation suggested that the module number should be multiplied for a quantitative phenotype [13]. Thus, the entire genetic effect of related modifier variants were substituted by a uniform random number to mimic the stochastic effect of a module. If there is only one module in the real phenotype, the simulation is too simple and the stochastic effect of each modifier variant should be simulated. In the formula, m represents the number of related modules and the additive stochastic component is the summation of the related module values. To be consistent with the data reported for autism and schizophrenia [10, 19, 20], hundreds of modifier variants should be assumed in this simulation. Because the number of modifier variants in a module may be from several to dozens [11], m was ranged from 1 to 50. The paternal phenotypic value (Xp), the maternal phenotypic value (Xm), and the additive stochastic component (the sigma component) have coefficients (a, b, and c, respectively) for their contributions. The coefficient c has a function to adjust the proportional contribution for the parental components and the additive stochastic component. The phenotypic value of a member of a generation was automatically calculated in a cell of by the spread sheet application using assigned Visual Basic assembly macro programs, and a population named generation zero ($G0$) was provided using only the sigma component for the calculation of the first generation ($G1$). As an example of the Visual Basic macros, macro a100 for a changeability simulation is shown in the Appendix. Although a traditional regression model (linear mixed-effects) has both fixed and random effects components, the components of the equation employed were all interpreted as random-effects and the model did not predict unknown continuous variables but rather the phenotypic diversity underpinned by a hidden genetic architecture. The mean value of the sigma component in a generation population is zero and the distribution approximates the normal distribution according to the central limit theorem [13]. Both the parental phenotypic values were calculated using the same formula as Xo in the previous generation, and the origins of the values were sigma components in $G0$ as described above. Therefore, when the distributional robustness was tested, the distribution of each generation could be approximated to the normal distribution in simulations. When distributional drifts or changeability could be simulated, the mean values for Xo, Xp, and Xm should be affected and changed. Given that phenotypic robustness could be demonstrated, the mean values for Xo, Xp, Xm, as well as the sigma component should be zero as well.

2.1. *Robustness simulations*

The male to female ratio in each generation was fixed to 1.0 and the marriage rate was fixed to 100%. Sex of an individual was determined randomly. Mating was done only within a generation and was also random, i. e. without positive or negative assortative mating. The random mating was simulated only within each generation and inter-generational mating did not occur. The parental phenotypic values (Xp and Xm) were carried over from the previous generation and the simulation of $G1$ was conducted employing $G0$ as described above. The offspring number for each couple was fixed at two for robustness simulations, resulting in a stable population size. Intrinsic distributional robustness and the phenotypic drift through generations was evaluated for different population sizes (10, 50, 100, 500, and 1,000), module numbers (m: 1, 2, 4, 10, 30, and 50), and coefficients (a, b, and c). To detect size-dependent drifts of generational populations, the population size was intentionally altered from unrealistic numbers for these simplified artificial simulations. To determine the optimum population size for the following changeability simulations, the smallest population size, for which distributional robustness could be obtained, was investigated. Automatically simulated generations were from $G1$ to $G100$. To assess the shape of the population, a generational population was depicted as a box-and-whisker diagram with outliers (small circles) whose values were between 1.5 and 3.0 times the box range and extremes (asterisks) whose values were more than 3.0 times the box range (IBM SPSS statistics, version 23). Each simulation was represented by a sequence of these boxplots.

2.2. *Changeability simulations*

Phenotypic changeability was evaluated on the basis of the robustness simulations. Sufficient population size ($n = 1,000$), for which distributional robustness could be maintained in the robustness simulations, was employed was employed for the changeability simulations. The extreme individuals with the lowest phenotypic values in a generation are supposed to be unreproductive. Their ability or opportunity in the mating arena is seriously reduced and they cannot leave offspring. Xo values of the extreme cases (0.2, 1, 2, 5, and 10%) were first eliminated from the spread sheet table as unreproductive individuals and residual reproductive individuals 998, 990, 980, 950, and 900 were all respectively randomly coupled within the generation as described above. The male to female ratio was fixed to 1.0 and the sex of an individual was determined randomly. To stabilize population size, hyper-reproductive couples, 1, 5, 10, 25, and 50 couples, respectively, were randomly selected regardless of the Xo values, and the selected couples were regarded as competent to have four offspring for the next generation. On the other hand, the other normal-reproductive couples (498, 490, 480, 450, and 400 couples) have 2 children. There were from 1 to 100 generations simulated. A generation population was depicted as the mean ± one standard deviation style (IBM SPSS statistics, version 23) and each simulation was represented by a line graph.

2.3. *Normality of the distribution and statistical analysis*

Descriptive calculations and statistical analyses were performed using a statistical software package (IBM SPSS statistics, version 23). The distribution of a generation is approximated to normal distribution in the artificial robustness simulations [13]. In order to access the intrinsic compatibility between distributional robustness and phenotypic changeability in complex conditions, the normality of the population distribution in a generation was evaluated using two methods for both the robustness and the changeability simulations. In descriptive statistics, a generation population with an absolute value for skewness and/or kurtosis that exceeded 2.0 was considered not to be normally distributed. p values based on the Shapiro-Wilk regression test statistic under the null hypothesis of normality were determined for the population distribution of a generation. When the p value was less than 0.05, the null hypothesis was rejected and the generation was considered not to be normally distributed.

3. Results

The simplified model, which was designed to be consistent with the correlation of parents and the *de novo* update of genetic identity in the zygote, demonstrated some pronounced functions of the stochastic epistasis component, as described below. Because preliminary simulations revealed that the phenotypic diversity of the generation population cannot be maintained without the stochastic epistatic component and the stability of the mean phenotypic values is conditional on the coefficients of the formula, to evaluate the influences of population size and the effect of the module number, the coefficient values were fixed ($a = 0.5$, $b = 0.5$, and $c = 1.0$). On the other hand, to evaluate the coefficient conditions, the module number was fixed to $m = 10$.

3.1. *Population size and distributional robustness*

Simulations were repeated five times and it was clearly demonstrated that the median value for each generation in small-sized simulations ($n = 10$ and 50) drifted randomly with the direction of the phenotypic change randomly determined (Fig. 1 gives a representative simulation for each population size). The phenotypic variance of the generation population also varies randomly in small-sized simulations. This phenotypic drift in small-sized simulation ($n = 10$) was confirmed for different module numbers ($m = 10$ in Fig. 1 and m = 1, 2, and 4 data not shown), and it was suggested that the drift can be exaggerated when either module number or stochasticity were increased. In marked contrast to the small-sized simulations, the simulation of $n = 500$ showed a significant distributional robustness from $G1$ to $G100$. This intrinsic robustness was confirmed by multiple simulations (five times) and ascertainment of the mean values. In this simulation ($n = 500$), the median or mean phenotypic value was almost the same through generations, while phenotypic diversity with extreme cases can also be maintained through generations. From these results, the population size for the following simulations were fixed to n = 1,000 to exclude the contamination from population size effects.

3.2. *Contributions of each component to the stochastic epistasis model*

Following exhaustive evaluation of every possibility, the simulations for each coefficient condition were repeated five times and three representative simulations are shown in Fig. 2. It was revealed that there are four conditions concerning the intrinsic distributional robustness. In the first condition ($a + b \leq 1$ and $c = 0$), although the phenotypic diversity could not be maintained without the additive stochastic epistasis component, the median value of the phenotype was the same as obtained for G1 across generations. In the second condition ($a + b \leqq 1$, $a + b > |c|$, and $c \neq 0$), the phenotypic diversity if $G1$ could not be maintained and the distribution range of the gen-
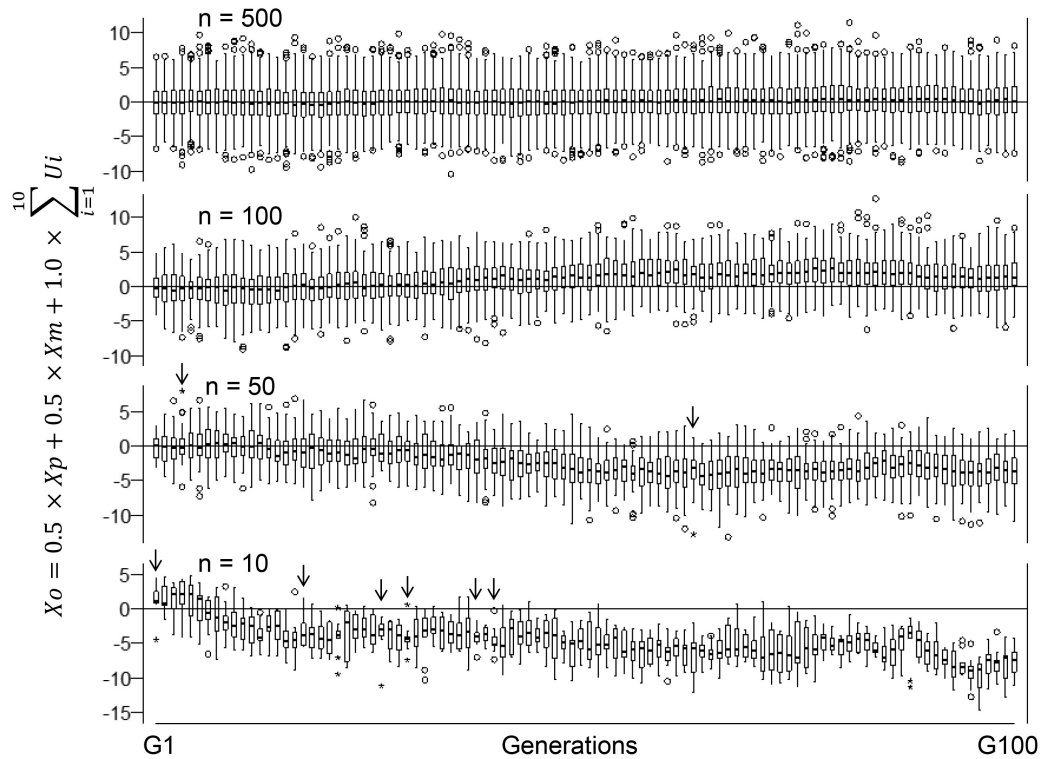
Fig. 1. Population size-dependent stability of the simulated phenotypic diversity. A simulation is illustrated as a sequence of boxplot diagrams. The phenotypic values (Xo) were automatically calculated using the formula (y-axis title) for each generation ($G1 - G100$) as described in Methods. Simulations were repeated five times with the population size varying from 10 to 500, and a representative simulation for each population size is shown. Small arrow heads indicate generations whose population was not normally distributed by an assessment of the absolute value for skewness and/or kurtosis ($\geqq 2.0$). The module number was fixed ($m = 10$) for these representative simulations.

eration population was gradually attenuated during the first several generations. After attenuation, the phenotypic diversity stabilized to a modest value. The median value of the phenotype was the same as $G1$ across generations. For the third condition ($a+b \leqq 1$, $a+b \leqq |c|$, and $c \neq 0$), both the phenotypic median and phenotypic variance could remain in this condition across generations. Therefore, in this condition, the perfect intrinsic robustness of phenotypic diversity could be obtained. In the fourth condition ($a + b > 1$ and $c \neq 0$), where the sum total of the parent's contribution coefficients (a and b) exceed 1.0, the phenotypic distribution suddenly deviated from approximately $G50$, although the direction and degree of the deviation was random, and the direction did not change in the simulation. Even for the deviated generations, the phenotypic diversity could be maintained as shown in Fig. 2.

3.3. Reproductive selection and phenotypic changeability

According to the percentage of the non-reproductive individuals whose phenotypic value was lowest in the generatiing population, the phenotypic distribution could be changed across generations (Fig. 3). The change occurred in the opposite direction to the non-reproductive extreme tail and the degree of change was determined by the value of coefficient c (data not shown). Together with the results of simulations with different numbers of stochastic epistatic modules

($m = 10$, 30, and 50), it was suggested that changeability could be exaggerated by an increase in the diversity range of the generation population when the module number increases or the stochasticity grows. The normality of the changing generating population could be maintained across generations. Although a continuous selection pressure generates a continual changeability across generations, a cessation of the pressure leads to a reappearance of distributional robustness and its eventual stabilization of the mean level.

3.4. Normality of the generation population

It is well known that the total sum of multiple uniform random numbers varies randomly, yielding a normally distributed histogram, and the histograms of stochastic epistasis models have previously been evaluated [13]. The normality of the population of each generation was assessed using descriptive statistics (the absolute value for skewness and/or kurtosis) and the Shapiro-Wilk regression test. It is only in simulations with small populations ($n = 10$ and 50), that cases with an absolute value for skewness and/or kurtosis exceeding 2.0 can be observed (Fig. 1). In almost all simulations the Shapiro-Wilk regression test detected some generations whose population may not be normally distributed and that exhibited a minimum correlation with the skewness/kurtosis results. In the simulations where population size was 10, 50, 100, 500, and 1,000 ($a = 0.5$, $b = 0.5$, $c = 1.0$,
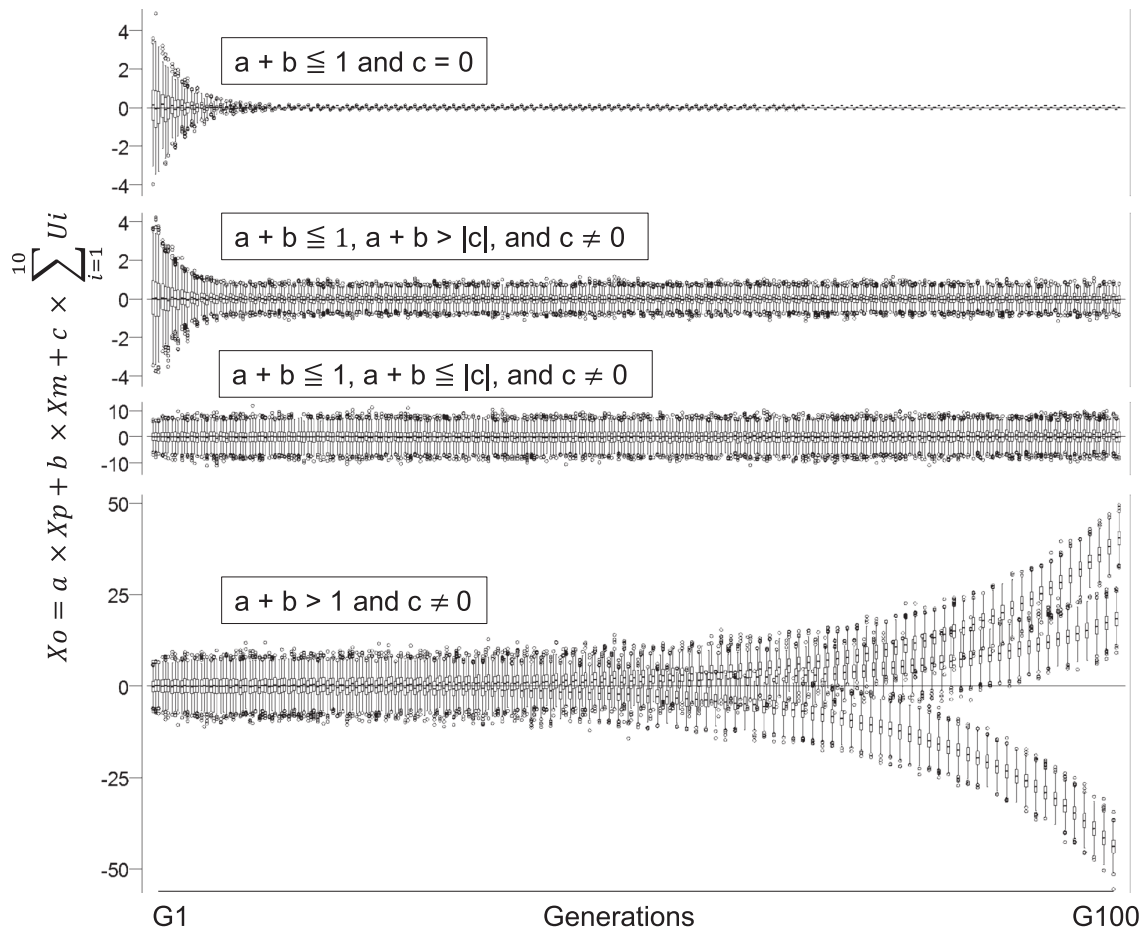
Fig. 2. Conditional effects of each coefficient of the formula (y-axis title). A simulation is illustrated as a sequence of boxplot diagrams. The phenotypic values (Xo) were automatically calculated using the formula for each generation ($G1 - G100$) as described in Methods. Simulations for each condition were repeated five times and three representative simulations are given for each condition (300 boxplots per condition) with $a = 0.5$, $b = 0.5$, and $c = 0$ for the condition $a + b \leqq 1$ and $c = 0$, $a = 0.5$, $b = 0.5$, and $c = 0.1$ for the condition $a + b \leqq 1$, $a + b > |c|$, and $c \neq 0$, $a = 0.5$, $b = 0.5$, and $c = 1.0$ for the condition $a + b \leqq 1$, $a + b \leqq |c|$, and $c \neq 0$, and $a = 0.53$, $b = 0.53$, and $c = 1.0$ for the condition $a + b > 1$ and $c \neq 0$, respectively. To exclude the contamination of the population size effect, the population size was fixed ($n = 1, 000$). The module number was also fixed ($m = 10$).

$m = 10$), the number of generations whose population was suggested to be not normally distributed by Shapiro-Wilk regression test ($p < 0.05$) was 4.4 ± 1.8, 6.8 ± 2.8, 6.4 ± 2.8, 6.0 ± 2.3, and 5.0 ± 2.6, respectively in 100 generations (each $n = 5$). In the representative simulations where population size was 10, 50, 100, 500, and 1,000, the maximal absolute values for skewness and/or kurtosis were 6.91, 3.21, 1.59, 0.58, and 0.35, respectively. In the changeability simulations with population size $n = 1,000$, the descriptive analyses revealed that the normality of the changing generation population could be preserved across generations. The normality was confirmed by maximal absolute values for skewness and/or kurtosis < 2.0. Under conditions where the percentage of nonreproductive individuals was 0.2, 1, 2, 5, and 10%, the maximal absolute value for skewness and/or kurtosis was: 0.35, 0.41, 0.39, 0.39, and 0.33, respectively, for the $m = 10$ simulations; 0.38, 0.35, 0.35, 0.47, and 0.38, respectively, for the $m = 30$ simulations; and 0.44, 0.40, 0.38, 0.46, and 0.41, respectively, for the $m = 50$ simulations.

4. Discussion

In the simulations to evaluate population size effects, it was demonstrated that small sized simulations ($n = 10$ or 50) can generate a phenotypic instability (drift) of the population distribution (Fig. 1). The median or phenotypic mean of each generation drifts randomly and the direction of the phenotypic change is determined by chance. Although the scale of the population size is different between mathematical simulations and real populations, the phenotypic drift has phenomenological similarities to the evolutionary 'genetic drift', which can cause a fitness-independent skewness of the pooled gene repertoire in small populations as an important driver of evolution [21]. Because the drift of the phenotypic distribution in the small sized simulations studied here provides a fitness-independent failure of neutrality with loss of genetic complexity, if the model is credible, the drifted population may consequently be exposed to selection pressures in the real world and the phenotypic differentiation among small populations may drive local adaptation similarly to the known 'genetic drift'-associated phenotypic changes [22]. Such fitness-independency associated with population size bottlenecking occurs easily in the stochastic epistasis model, in which the link between
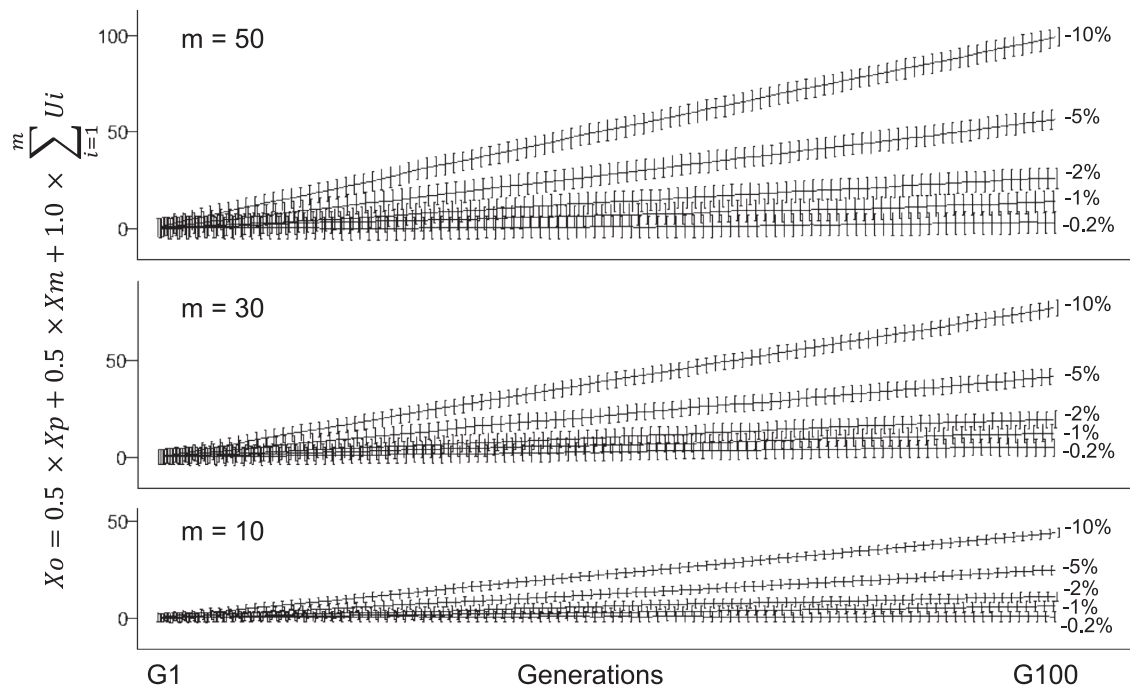
Fig. 3. Phenotypic changeability in a selection pressure where the lowest extremes in the population cannot leave offspring. A simulation is illustrated as a line graph (the mean value ± one standard deviation). The phenotypic values (Xo) were automatically calculated using the formula (y-axis title) for each generation ($G1 - G100$) as described in Methods. The percentage of nonreproductive extreme cases is shown at the right end of the simulation (from 0.2% to 10% with a minus symbol). To exclude the contamination of population size effect, the population size was fixed ($n = 1,000$).

related modifier variant and phenotypic outcome is hidden in a large population. The position of a modifier-carrier can be anywhere in the phenotypic distribution and the size of any research sampling in the population may be too bottlenecking (too small) for the modifier carriers in sample to enjoy the neutrality of the phenotypic mean [11]. Therefore, in the stochastic epistasis concept, the fitness-independent failure of neutrality might be common in the samplings for genetic research and some of the related modifier variants in the sample could be detected by a usual study sampling by chance. However, it is supposed that the associations detected between complex conditions and modifiers can hardly be replicated by additional research, the modifiers are not necessarily detected only in the extreme cases, and the presence of variants may only be revealed in a few of the extreme cases. From this perspective, another sampling may result in the detection of another association. In actual research on complex human conditions including autism and schizophrenia, hundreds of candidate variants have been found and most of these variants explain only a modest amount of the observed heritability (the majority of the variants assumed are missing) [10, 19, 20]. The reported variants including *de novo* mutations can be detected in a minor number of cases and sometimes show incomplete penetrance in relatives and low result replication rates in different study samples [12]. These similarities between the mathematical simulations and complex human conditions may increase the credibility of the stochastic epistasis model for complex conditions.

In contrast to the phenotypic instability in small-sized simulations of the model, there was a pronounced intrinsic robustness of the phenotypic diversity in large-sized simulations ($n = 500$ or 1000) with an optimal condition ($a+b \leq 1$, $a+b \leq |c|$, and $c \neq 0$) in which the phenotypic median (or mean) can be kept neutral (value zero)

even until the 100th generation (G100) (Fig. 2). Normality of the generation distribution was perfectly maintained in the simulations as confirmed by the maximal absolute value for skewness and/or kurtosis (< 2.0). When the additive stochastic epistasis component has no effect ($a+b \leq 1$ and $c \neq 0$) or less effect than the parental correlative components ($a+b \leq 1$, $a+b > |c|$, and $c \neq 0$), the median value of each generation can be constant with no or modest diversity. Therefore, the stochastic epistasis component is critical for the framework of the distributional robustness of phenotypic diversity, and the normal shape with gently-sloping bilateral tails of the population distribution depends on the stochastic epistatic contribution. The fitness-independent maintenance of phenotypic median (or mean) without wide-ranged diversity ($a+b \leq 1$, $a+b > |c|$, and $c \neq 0$) may be utilized for the simulation of stabilizing selection.

In the changeability simulations, the same formula as the robustness simulations showed a complexity-dependent phenotypic changeability in a selection situation where individuals with the lowest Xo values are nonreproductive (Fig. 3). In the same way as the fitness-independent failure of the fitness neutrality in size-bottlenecking populations, the fitness-dependent phenotypic change can be exaggerated by the increase of the stochastic epistatic contribution. Because both the fitness-independent and -dependent phenotypic changes are regarded as the core of local adaptation and evolution [21–23], the implications of these mathematical phenotypic changes for complex human conditions should be discussed as follows. There is no heritability without genetic diversity [24], and co-segregation between a phenotypic characteristic and the related gene variants is one of the essential prerequisites for differentiation, development, disease susceptibility, evolutionary adaptation, and speciation. In

a fitness-associated complex condition whose normal distribution includes the majority around the phenotypic mean and the flanking bilateral extreme tails, if the co-segregation is fixed beyond generations, the distribution has a risk of being skewed by the burden of selection process in which an extreme tail is nonreproductive. Why can the variant-based normally distributed phenotypic diversity be maintained through generations? Why can the nonreproductive tail (e.g. autism and schizophrenia) of human diversity remain prevalent [9, 25]? Fitness-independent mechanisms, temporal or environmental fitness changes or trade-offs, population benefit theory, and the combinations, as a homeostatic mechanism have been repeatedly addressed as the solutions [23, 26–33]. In a novel concept, the stochastic epistasis perspective, the difference from the temporal fitness trade-offs in balancing selection (the classical neutral theory of molecular evolution) is the stochasticity or unpredictability of the fitness oscillations. Importantly, in the novel model, the carriers of modifier variants are ubiquitously scattered in the entire population distribution [9, 11–13]. In the mathematical model which was used for simulations here, the reproductive majority of a generation population can retain the genetic mechanism which regenerates the extreme in the next generation. Therefore, phenotypic change is available without damage to the genetic mechanism for the extreme cases in the stochastic epistatic model. Although only surviving individuals can be involved in adaptation and evolutionary changes on the cost of the nonreproductive extremes, the finding that phenotypic changes are exaggerated by increased stochasticity or complexity in the extreme-regeneration mechanism may emphasize the implication of the presence of nonreproductive extreme cases in adaptation and quantitative population change. Because the increase of stochasticity or complexity in the extreme-regeneration mechanism is a manifestation of the intensification of the stochastic epistasis mechanism and the phenotypic change is based on the fitness benefits, the fitness benefit of the population depends on the stochastic epistasis and the regeneration of nonreproductive individuals. The presence of the nonreproductive tail is not for the regeneration of the extreme tail, but for the reproductive majority's competence to change, suggesting that the extremes repeatedly exist for others [9]. Furthermore, the evolution of the stochasticity or complexity of the extreme-regeneration (stochastic epistasis) mechanism may be referred to as the evolution of changeability. Each of the modifier variants is indeed neither necessary nor sufficient for the establishment of the extreme cases in the stochastic epistasis model, similarly to the quantitative trait loci (QTLs) framework [1], but complex conditions with huge phenotypic diversity like autistic assets may be characterized by the increased number of the related stochastic epistatic modules. A huge hidden genetic variation in the stochastic epistasis modules may be an attractive candidate for the signature of long-lasting historical quantitative changes as well as an index of evolutionary potential [34–36].

5. Conclusions

The stochastic epistasis concept is a new neutral theory of quantitative changes. In contrast to the traditional theory with contextual prediction or explanation for the temporal change of phenotypic fitness [23, 28, 32], it had been specified previously that the phenotypic oscillation of stochastic epistasis is random or unpredictable [9, 11–13]. In this approach, the stochastic effects of related epistatic modifier variants make the network's effect stochastic and multiple related

networks (modules) stochastically affect a phenotypic outcome [11]. The stochastic epistatic module can be a reservoir of genetic variants and an important role of the genetic diversity is the maintenance of the stochasticity. The hidden genetic variation can also fuel evolution when circumstances change through gene-environment interactions [36]. The stochasticity hides related modifier variants and modules, but heritability of the phenotype can be preserved in any portion of the phenotypic distribution. Stochastic epistasis-associated conditions are characterized by a high concordance in monozygotic twins and a low concordance in dizygotic twins and recognized to be highly heritable but not inherited.

Modest correlations between core characteristics in a complex trait and the sizable overlap among multiple distinct complex conditions can be illustrated by the infidelity of pleiotropy in stochastic epistasis. Additionally, here it was demonstrated that both the intrinsic distributional robustness of phenotypic diversity and quantitative changeability can be mathematically explored with a stochastic epistasis model, and the fast and strong plasticity of the nonreproductive extreme tail may be essential for both survival and population changes under selection pressures. Stochastic epistasis may also be the critical mechanism for evolvability. After a population size bottleneck episode, the loss of genetic complexity in the model can be a cue for fitness-independent phenotypic changes and local adaptation. Furthermore, the model explains the research situation where the result of an initial genetic study cannot be replicated by subsequent studies and a large-scale study or meta-analysis of multiple studies is unable to screen the candidate genes. Further simulation to illustrate how extreme individuals can stay at low fitness for generations may be warranted to further understand this model of stochastic epistasis.

Appendix

An example of the Visual Basic macros (Microsoft Excel 2010) for a changeability simulation. The worksheet G0 for generation $G0$ and the worksheet m for an integer m are prepared in advance. The row for generation $G0$ with 10 modules is Q in the $G0$ worksheet. This macro ($a100$) should be run when a new worksheet is active.

```
Sub a100()
Range("A1:A1000") = Worksheets("G0").Range("Q1:Q1000").Value
Dim m As Integer
For m = 1 To 100
Worksheets("m").Cells(1, 1) = m
Call copy01
Call matching01
Call delete01
Call randomize01
Call copy02
Call hyperreproductive01
Call kids01
Call copy03
Call x01
Next m
End Sub
 Private Sub copy01()
m = Worksheets("m").Cells(1, 1)
Range(Cells(1, 5 * m − 2), Cells(1000, 5 * m − 2)) = Range(Cells(1, 5 * m
        − 4), Cells(1000, 5 * m − 4)).Value
End Sub
 Private Sub matching01()
m = Worksheets("m").Cells(1, 1)
Range(Cells(1, 5 * m − 2), Cells(1000, 5 * m − 2)) _
        .Sort Key1:=Range(Cells(1, 5 * m − 2), Cells(1000, 5 * m − 2)),
            Order1:=xlAscending
```

```
End Sub
 Private Sub delete01 ()
m = Worksheets("m").Cells (1, 1)
Range(Cells(1, 5 * m − 2), Cells(50, 5 * m − 2)).Delete Shift :=xlShiftUp
End Sub
 Private Sub randomize01()
m = Worksheets("m").Cells (1, 1)
Dim i As Integer
For i = 1 To 950
    Cells (i, 5 * m − 1) = Rnd()
Next i
Range(Cells(1, 5 * m − 2), Cells(950, 5 * m − 1)) _
    .Sort Key1:=Range(Cells(1, 5 * m − 1), Cells(950, 5 * m − 1)), Order1
        :=xlAscending
Range(Cells(1, 5 * m − 1), Cells(1000, 5 * m − 1)).Clear
End Sub
 Private Sub copy02()
m = Worksheets("m").Cells (1, 1)
Range(Cells(476, 5 * m − 2), Cells(950, 5 * m − 2)).Copy
Range(Cells(1, 5 * m − 1), Cells (475, 5 * m − 1)). PasteSpecial
Range(Cells(476, 5 * m − 2), Cells(950, 5 * m − 2)).Delete Shift :=
        xlShiftUp
End Sub
 Private Sub hyperreproductive01 ()
m = Worksheets("m").Cells (1, 1)
Range(Cells(451, 5 * m − 2), Cells(475, 5 * m − 1)).Copy
Range(Cells(476, 5 * m − 2), Cells(500, 5 * m − 1)). PasteSpecial
End Sub
 Private Sub kids01 ()
m = Worksheets("m").Cells (1, 1)
Range(Cells(1, 5 * m − 2), Cells(500, 5 * m − 1)).Copy
Range(Cells(501, 5 * m − 2), Cells(1000, 5 * m − 1)). PasteSpecial
End Sub
 Private Sub copy03()
m = Worksheets("m").Cells (1, 1)
Range(Cells(1, 5 * m), Cells(1000, 5 * m)) = Worksheets("G0").Range("Q1:
        Q1000").Value
End Sub
 Private Sub x01()
m = Worksheets("m").Cells (1, 1)
Dim i As Integer
For i = 1 To 1000
    Cells (i, 5 * m + 1) = 0.5 * Cells (i, 5 * m − 2) + 0.5 * Cells (i, 5 *
        m − 1) + 1 * Cells (i, 5 * m)
Next i
End Sub
```

Acknowledgments

Conflict of Interest

All authors declare no conflicts of interest.

References

[1] Plomin R, Owen MJ, McGuffin P (1994) The genetic basis of complex human behaviors. *Science* **264(5166)**, 1733-1739.

[2] Plomin R, DeFries JC, Knopik VS, Neiderhiser JM (2016) Top 10 replicated findings from behavioral genetics. *Perspectives on Psychological Science* **11(1)**, 3-23.

[3] Fischer B, Taborsky B, Dieckmann U (2009) Unexpected patterns of plastic energy allocation in stochastic environments. *The American Naturalist* **173(3)**, e108-e120.

[4] Raser JM, O'shea EK (2005) Noise in gene expression: origins, consequences, and control. *Science* **309(5743)**, 2010-2013.

[5] Shin T, Kraemer D, Pryor J, Liu L, Rugila J, Howe L, Buck S, Murphy K, Lyons L, Westhusin M (2002) Cell biology: a cat cloned by nuclear transplantation. *Nature* **415(6874)**, 859.

[6] Lark KG, Chase K, Adler F, Mansur LM, Orf JH (1995) Interactions between quantitative trait loci in soybean in which trait variation at one locus is conditional upon a specific allele at another. *Proceedings of the National Academy of Sciences of the United States of America* **92(10)**, 4656-4660.

[7] Kroymann J, Mitchell-Olds T (2005) Epistasis and balanced polymorphism influencing complex trait variation. *Nature* **435(7038)**, 95-98.

[8] Pettersson M, Besnier F, Siegel PB, Carlborg Ö (2011) Replication and explorations of high-order epistasis using a large advanced intercross line pedigree. *Plos Genetics* **7(7)**, e1002180.

[9] Ijichi S, Ijichi N, Ijichi Y, Kawamura Y, Hashiguchi T, Morioka H (2008) For others: Epistasis and the evolutionary survival of an extreme tail of the quantitative distribution of autistic assets. *Medical Hypotheses* **70(3)**, 515-521.

[10] Beaudet AL (2007) Autism: highly heritable but not inherited. *Nature Medicine* **13(5)**, 534-536.

[11] Ijichi S, Ijichi N, Ijichi Y, Nagata J, Imamura C, Sameshima H, Kawaike Y, Morioka H (2015) The origin of population diversity: Stochastic interactions between a modifier variant and the individual genetic background. *Natural Science* **7(5)**, 255-265.

[12] Ijichi S, Ijichi N, Ijichi Y, Sameshima H, Morioka H (2011) The genetic basis of phenotypic diversity: autism as an extreme tail of a complex dimensional trait. In, Amaral D et al.(Eds.) *Autism Spectrum Disorders: The Role of Genetics in Diagnosis and Treatment*. New York, Oxford University Press.

[13] Ijichi S, Ijichi N, Ijichi Y, Sameshima H, Saeki Y, Hall W, Morioka H (2011) Non-additive interactions between monomorphic and polymorphic loci: a theoretical explanation of the genetic underpinning for complex traits. *Advances in Medicine and Biology* **13**, 133-148.

[14] Raser JM, O'shea EK (2004) Control of stochasticity in eukaryotic gene expression. *Science* **304(5678)**, 1811-1814.

[15] Ansel J, Bottin H, Rodriguez-Beltran C, Damon C, Nagarajan M, Fehrmann S, François J, Yvert G (2008) Cell-to-cell stochastic variation in gene expression is a complex genetic trait. *Plos Genetics* **4(4)**, e1000049.

[16] Chalancon G, Ravarani CN, Balaji S, Martinez-Arias A, Aravind L, Jothi R, Babu MM (2012) Interplay between gene expression noise and regulatory network architecture. *Trends in Genetics* **28(5)**, 221-232.

[17] Feinberg AP, Irizarry RA (2010) Stochastic epigenetic variation as a driving force of development, evolutionary adaptation, and disease. *Proceedings of the National Academy of Sciences of the United States of America* **107(1)**, 1757-1764.

[18] Magklara A, Lomvardas S (2013) Stochastic gene expression in mammals: lessons from olfaction. *Trends in Cell Biology* **23(9)**, 449-456.

[19] Lander ES (2011) Initial impact of the sequencing of the human genome. *Nature* **470(7333)**, 187-197.

[20] Van Dongen J, Boomsma DI (2013) The evolutionary paradox and the missing heritability of schizophrenia. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics* **162(2)**, 122-136.

[21] Taub DR, Page J (2016) Molecular signatures of natural selection for polymorphic genes of the human dopaminergic and serotonergic systems: a review. *Frontiers in Psychology* **7**, 857.

[22] Willi Y, Van Buskirk J, Schmid B, Fischer M (2007) Genetic isolation of fragmented populations is exacerbated by drift and selection. *Journal of Evolutionary Biology* **20(2)**, 534-542.

[23] Nesse RM, Williams GC (1996) *Why We Get Sick: The New Science of Darwinian Medicine*. New York, Vintage.

[24] Lynch M, Walsh B (1998) *Genetics and Analysis of Quantitative Traits*. Sunderland, MA: Sinauer Associates, Inc..

[25] Caspermeyer J (2015) An evolutionary approach reveals new clues toward understanding the roots of schizophrenia. *Molecular Biology and Evolution* **32(6)**, 1657, 1658.

[26] Baron-Cohen S (2002) The extreme male brain theory of autism. *Trends in Cognitive Sciences* **6(6)**, 248-254.

[27] Fitzgerald M (2004) Autism and Creativity: Is there a Link between Autism in Men and Exceptional Ability? *British Medical Journal* **328(7448)**, 1139.

[28] Keller MC, Miller G (2006) Resolving the paradox of common, harmful, heritable mental disorders: which evolutionary genetic models work best? *Behavioral and Brain Sciences* **29(4)**, 385-404.

[29] Badcock C, Crespi B (2006) Imbalanced genomic imprinting in brain development: an evolutionary basis for the aetiology of autism. *Journal of Evolutionary Biology* **19(4)**, 1007-1032.

[30] Pearlson GD, Folley BS (2007) Schizophrenia, psychiatric genetics, and Darwinian psychiatry: an evolutionary framework. *Schizophrenia Bulletin* **34(4)**, 722-733.

[31] Reser JE (2011) Conceptualizing the autism spectrum in terms of natural selection and behavioral ecology: the solitary forager hypothesis. *Evolutionary Psychology* **9(2)**, 207-238.

[32] Dudley JT, Kim Y, Liu L, Markov GJ, Gerold K, Chen R, Butte AJ, Kumar S (2012) Human genomic disease variants: a neutral evolutionary explanation. *Genome Research* **22(8)**, 1383–1394.

[33] Xu K, Schadt EE, Pollard KS, Roussos P, Dudley JT (2015) Genomic and network patterns of schizophrenia genetic variation in human evolutionary accelerated regions. *Molecular Biology and Evolution* **32(5)**, 1148-1160.

[34] Gibson G, Dworkin I (2004) Uncovering cryptic genetic variation. *Nature Reviews Genetics* **5(9)**, 681-690.

[35] Breen MS, Kemena C, Vlasov PK, Notredame C, Kondrashov FA (2012) Epistasis as the primary factor in molecular evolution. *Nature* **490(7421)**, 535-538.

[36] Paaby AB, Rockman MV (2014) Cryptic genetic variation: evolution's hidden substrate.*Nature Reviews Genetics* **15(4)**, 247-258.