Finding Proteins Whose Expression Levels Depend on Bloodline in Wagyu

Takatoshi Fujiki
Graduate School of Systems Engineering,
Wakayama University, Japan
Email: s101044@sys.wakayama-u.ac.jp

Satoshi Sakaguchi
Graduate School of Systems Engineering,
Wakayama University, Japan
Email: s121016@sys.wakayama-u.ac.jp

Takuya Yoshihiro
Faculty of Systems Engineering,
Wakayama University, Japan
Email: tac@sys.wakayama-u.ac.jp

***Abstract*—Wagyu is known as beef brand with marbling character, in which the lineage of sires is significantly important to provide quality beef continuously. Sires of Wagyu have been improved through the dedicated efforts of inbreeding, to obtain excellent genetic ability to yield quality beef. In this decade, rapid growth of the technologies to analyze genes and proteins brings us a chance to improve the quality of beef using more direct and precise tools and knowledge. Tremendous amount of relations among genes, proteins, and traits have been clarified, and the knowledge can be potentially utilized to improve the quality of beef. However, there is scarcely a method that analyze bloodlines of sires and cattle to connect bloodline to genes and proteins. In this paper, we newly propose a method to treat bloodline of livestock animals on computers, and to find proteins whose expression levels have strongly related to the bloodline of cattle. With the proposed method, we firstly have a mean to know the relation between bloodline and proteins.**

Keywords*—*Beef Brand; Bloodline; Protein Expression.

I. INTRODUCTION

Wagyu is Japanese native beef cattle known for marbling character, in which the linage of sires is significantly important to provide quality beef. Sires of Wagyu have been improved through the dedicated efforts of inbreeding, to obtain excellent genetic ability to yield quality beef. Currently, by using frozen sperms, quality beef cattle have been produced continuously from excellent genes of Wagyu sires. The lineage of sires, which guarantees the quality of Wagyu, is the precious genetic source that is essential to yield quality beef continuously.

For breeders of Wagyu cattle and sires, selecting sires (i.e., selecting sperms) for a new born cattle and a sire is one of the most important tasks, because the genetic character (hereafter, we call it the *lineage*) of a new born cattle and a sire is deeply related to the quality of beef yielded by them. Thus, traditionally, breeders utilize the values so called *breeding values*, which expresses the genetic ability of sires to produce the quality beef cattle, in selecting sires to use. In general, the breeding values are calculated using the BLUP (Best Linear Unbiased Prediction) method [1]. The BLUP method is a statistical prediction method to estimate the breeding values from the past results of beef grades of the sire's children, the bloodline information, and so on. Breeders of Wagyu have improved the genetic ability of sires for a long time by selecting good sires via breeding values to produce excellent descendants.

On the other hand, recently, many mechanisms of various life phenomena have been clarified due to the improvement of the technology to analyze genes and proteins of samples. For example, the techniques so-called microarray and 2D-electrophoresis enable us to measure expression levels of thousands of genes and proteins simultaneously [2]. With these high-throughput experimental methods, many specific mechanisms of creatures have been clarified so far. If the mechanism to yield quality beef is clarified, i.e., if the mechanism to connect both (i) from the bloodline to proteins, and (ii) from proteins to the beef quality, are clarified, a new and more efficient methodology to improve beef quality may be developed.

As for (ii), i.e., on the protein-protein interaction or the protein-phenotype relation, there are so many studies proposed so far. Bayesian networks [3] construct an interaction network model among proteins and phenotypes based on conditional probability. As other methods, we developed an algorithm to predict interactions among three proteins A, B and C, based on correlation coefficient [4], and conditional probability [5]. If we regard C as a phenotype, this method can be used to investigate the relation between proteins and phenotypes. However, there are few method that investigates (i), i.e., the relation between bloodline and proteins.

In this paper, we propose a new method that investigates the relation between bloodline and proteins; specifically, we try to find proteins whose expression levels are controlled by the lineage of beef cattle. As for these proteins, if there are two cattle that lineage is similar, then the expression levels are also similar, and otherwise the expression levels are not necessarily similar. By finding such proteins, we can determine a set of proteins in order to investigate the mechanism to improve the quality of beef, as well as we can specify the genes included in the lineage of sires that are deeply related to the expression levels of these proteins. To the best of our knowledge, this is the first study that tries to investigate the relation between the lineage and proteins.

This paper is organized as follows: In Section 2, we describe the protein whose expression levels depend on bloodlines, in addition to explain the input data of the proposed algorithm. In Section 3, we describe the proposed algorithm in detail. In Section 4, we present the method and the result of the evaluation. Finally, in Section 5, we conclude the paper.

II. PROTEINS WHOSE EXPRESSION LEVELS DEPEND ON BLOODLINE

A. Introducing Lineage Vectors to Represent Bloodlines

The genetic characters of Wagyu have been improved for a long time through dedicated efforts on inbreeding to generate

ISBN: 978-1-61208-295-0

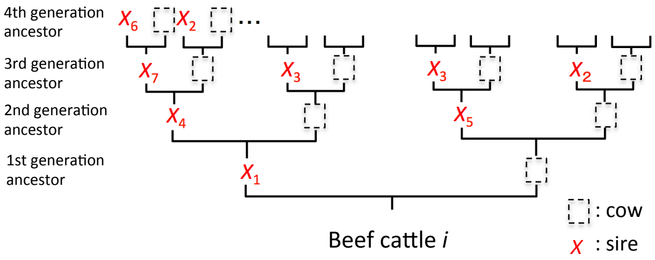

Figure 1. Family Tree of Cattle

excellent bloodlines of sires. All the ancestors of Wagyu cattle have excellent bloodlines, and the genetic character yields high-quality Wagyu beef. In general, the bloodline of each Wagyu cattle is recorded as a family tree that includes all ancestor sires over several generations. However, in order to treat bloodlines in computers, it is desirable to convert the family tree into a form with which we can easily treat it mathematically. So, we introduce a *lineage vector* of Wagyu cattle that expresses its genetic characters in a computable form.

Before the definition of lineage vectors, we first explain the records of family trees of Wagyu cattle. Figure 1 illustrates the family tree of one Wagyu cattle, where the root is the Wagyu cattle, and $X_1, X_2, \ldots, X_n$ are the ancestor sires of it. In general, cows are not included in the family tree because the effect of cows on genetic character is not strong; a sire can be the father of thousands of cattle whereas a cow can be the mother of no more than ten cattle. Note that a sire may appear more than twice in a family tree because Wagyu cattle as well as sires are generated from frozen sperms, and so the frozen sperms of an excellent sire tend to be used repeatedly even in different generations in the family tree.

Each Wagyu cattle has its family tree, and the lineage vector that is converted from the family tree. The line vector represents the ratio of genetic information inherited from each sire to the cattle. According to the genetic mechanism of inheritance, it is naturally assumed that the genetic information of cattle inherited from 1st generation (father) sire is 50%, and that from 2nd generation (grandfather) sire is 25%. Namely, the genetic information inherited from a k-th generation sire is $(\frac{1}{2})^k$. For instance, in Figure 1, 12.5% of the genetic information is inherited from the 3rd generation ancestors X_2, X_3, and X_7, and 6.25% from the 4th generation ancestors X_2 and X_6. If a sire appears more than twice in a family tree, the ratio of genetic information inherited is the sum of them, i.e., the genetic information inherited from X_2 is 12.5%+6.25%=18.75% in Figure 1.

Now, we define the lineage vector formally. Let $i (1 \leq i \leq b)$ be a cattle, $X_t (1 \leq t \leq n)$ be a sire, and $a_t^{(i)} (0 \leq a_t^{(i)} < 1)$ be the ratio of genetic information that cattle i is inherited from a sire X_t. Then, the lineage vector $B^{(i)}$ of cattle i is defined on the vector space where every possible sire has its corresponding basis, as follows:

$$B^{(i)} = \left(a_1^{(i)}, a_2^{(i)}, \cdots, a_t^{(i)}, \cdots, a_n^{(i)} \right) \quad (1 \leq i \leq b) \qquad (1)$$

B. Expression Profiles of Proteins

Recent rapid growth of biological technology enabled us to analyze proteins included in a tissue of creatures with low cost in short time. There are several technologies that analyze proteins: one of major approaches is to obtain expression profiles, which analyzes the amount of each proteins included in a tissue, and the 2-dimensional electrophoresis is the representative technique to obtain expression profiles efficiently [2]. In this paper, we assume the protein expression profile of Wagyu cattle as the input data of the proposed algorithm.

We let $P_j (1 \leq j \leq m)$ be a protein, and the expression profile as the input of our algorithm consists of the expression levels $e_{P_j}^{(i)}$ for every proteins P_j and beef cattle i. We assume that the expression profile is normalized properly with some normalization methods.

C. Proteins Whose Expression Levels Depend on Bloodlines

In this paper, we try to find proteins that the expression levels depend on the bloodline of cattle, i.e., the expression levels are significantly related to the bloodline. Note that, if we choose a protein that is significantly related to the bloodline, the expression levels of two cattle of similar bloodlines will take similar values, and otherwise the two expression levels will have no relation. The problem that we try to solve is to measure the strength of this tendency between expression levels and bloodline for each protein in the expression profile.

We show an example in Figure 2 and Figure 3. Figure 2 is the case where the expression levels are related to the bloodline. Here, there are four vicinities of bloodlines and several samples (i.e., cattle) belong to them. The variance of each vicinity compared to that of all samples takes relatively small value. On the other hand, Figure 3 is the case where there is no relation between expression levels and the bloodline. The variance of every vicinity is almost the same as the variance of all samples.

Let us consider the problem more specifically using the lineage vectors of cattle. The lineage vectors belong to the *lineage space*, which is n-dimensional Euclidean space. If we suppose a point p in the lineage space, we can define the vicinity of p as the set of points within the Euclidean distance of ϵ. All we have to do is to examine every different coordinate in the lineage space, and for each of the coordinates, compute the variance of the expression levels of the samples included in the vicinity. If we compute the average of all the computed variances for each protein, the average indicates the strength of the relation between (the expression levels of) the protein and bloodline.

Note that there arises a problem of computation time because the lineage space has very large dimension n. In the next section, we will present the algorithm based on Gaussian processing [6] to cope with this problem.

III. The Proposed Method

A. Algorithm Design

As described in the previous section, we can retrieve the protein whose expression level is controlled by bloodline by examining the variance at every coordinate in the lineage

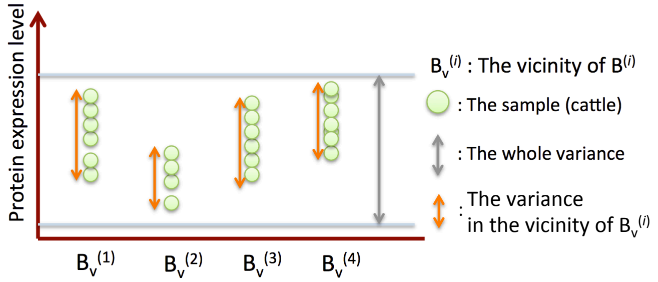

Figure 2. Protein Whose Expression Levels Depend on Bloodlines

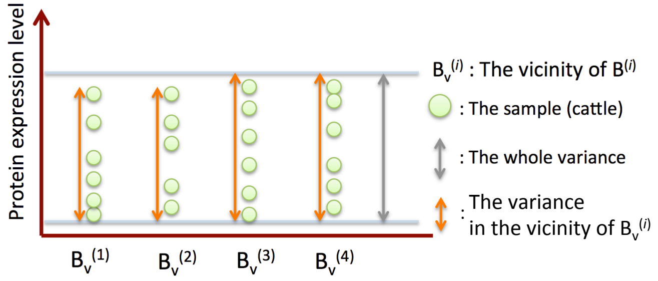

Figure 3. Protein Whose Expression Levels Don't Depend on Bloodlines

space. However, because the lineage space is continuous, to examine every possible coordinate in the lineage space is impossible. Moreover, because the dimension n of the lineage space is supposed as large as several hundred in a standard Wagyu dataset, the density of samples in the lineage space is very sparse even if the number of samples is several ten thousands. It is natural that the densities of samples at some coordinates are not large enough to guarantee statistical reliability of the computed variance values.

The main idea of the proposed algorithm is to partition the lineage space into many hypercubes (that we call *cells* hereafter) that have the same side length in every dimension, and we only examine the central coordinates of these cells. Note that the number of cells to be examined is tremendous because the lineage space has n dimensions. Thus, we only examine the cells to which at least one sample belongs to reduce the number of cells to examine. Furthermore, to guarantee the statistical reliability of computed variances, we calculate the variance of a cell only if the density of samples in the cell is larger than a threshold T.

Formally, the given parameter l, which represents the length of cell side, we define the coordinate of centers of cells in the n-dimensional lineage space as follows:

$$c = (c_1, c_2, \cdots, c_n), \tag{2}$$

where

$$c_t = ceiling\left(floor\left(\frac{a_t^{(i)}}{l/2}\right)/2\right) \cdot l \;\; (1 \leq t \leq n) \tag{3}$$

Then, the formal algorithm description is given in Table 1. In line 4, we compute the sample density at the center of each cell, and for each cells that densities are larger than T, we compute the variance of the cell in line 6. Finally in line 10, function compute_score() calculates the average of all the computed variances for each protein P_j as the *control score*, which represents the strength of the relation that the protein is controlled by bloodline. The lower the controlled score of a protein is, the stronger the expression levels of the protein are controlled by bloodline.

TABLE I. The Proposed Algorithm

```
1   foreach protein $P_j$ $(1 \leq j \leq m)$
2     foreach cell $c$
3       if samples exist in $c$
4         $Dens(c) \leftarrow$ compute_density()
5         if $D_c \geq T$
6           $Var(c, P_j) \leftarrow$ compute_variance()
7         end
8       end
9     end
10    $Score(P_j) \leftarrow$ compute_score()
11  end
```

The functions to estimate the sample density (i.e., compute_density()), the variance of expression levels (i.e., compute_variance()), and the control score (i.e., compute_score()) are described in the following Sections III-B, III-C, and III-D, respectively.

B. Estimating Sample Density

In this section, we describe the function compute_density() that appears in line 4 of the proposed algorithm, which is the function to calculate the density of the center of a cell. Let $c = (c_1, c_2, \cdots, c_n)$ be the coordinates of the center point at which we want to compute the density. We apply the Kernel density estimation [7] to estimate the density, which is a widely used non-parametric density estimation method.

In our density estimation function, we first calculate the distance from the center c to each sample in the cell. Next, we estimate the sample density at the center c by accumulating the density according to the distance using the Kernel density function. Note that we use the Euclidean distance in this method.

Formally, the distance between the center c of the cell and the sample point of the beef cattle i are defined as

$$Dist(i, c) = \sqrt{\sum_{t=1}^{n} (a_t^{(i)} - c_t)^2 \;\; (1 \leq t \leq n)} \tag{4}$$

Then, as the Kernel function $K(\frac{Dist(i,c)}{h})$, we use a general multi-dimensional Gaussian function, i.e.,

$$K\left(\frac{Dist(i, c)}{h}\right) = \frac{1}{(\sqrt{2\pi h^2})^n} exp\left(-\frac{1}{2h^2}(Dist(i, c))^2\right), \tag{5}$$

where h is a parameter that represents the bandwidth. As a result, the estimated density $Dens(c)$ of the center point c of the cell is given as follows:

$$Dens(c) = \frac{1}{b}\sum_{i=1}^{b} K\left(\frac{Dist(i, c)}{h}\right) \tag{6}$$

Copyright (c) IARIA, 2013. ISBN: 978-1-61208-295-0

C. Estimating Variance of Expression Levels

In this section, we describe the function compute_variance() that appears in line 6 of the proposed algorithm, which computes the variance of the expression levels at the center of a cell.

We estimate the variance of expression levels using a Gaussian process. Namely, we calculate the weighted variance of expression levels of samples using a weight function where the weight is determined according to the distance between the center c and the point of samples. Formally, the estimated variance $Var(c, P_j)$ of expression levels for protein P_j at the coordinate c is represented as

$$Var(c, P_j) = \frac{\sum_{i=1}^{b} \left(\left(e_{P_j}^{(i)} - Avg_j(c) \right)^2 K\left(\frac{Dist(i,c)}{h} \right) \right)}{\sum_{i=1}^{b} K\left(\frac{Dist(i,c)}{h} \right)} \quad (1 \leq j \leq m), \qquad (7)$$

where $e_{P_j}^{(i)}$ is the expression level of the protein P_j corresponding to the beef cattle i, and $Avg_j(c)$ is the average of the expression levels for protein P_j at c represented as follows:

$$Avg_j(c) = \frac{\sum_{i=1}^{b} \left(e_{P_j}^{(i)} K\left(\frac{Dist(i,c)}{h} \right) \right)}{\sum_{i=1}^{b} K\left(\frac{Dist(i,c)}{h} \right)} \qquad (8)$$

D. Calculation of Control Score

We describe the process to calculate the controlled score of each proteins. The controlled score $Score(P_j)$ of the protein P_j is the average of the estimated value which the variance of the expression level at the central point $c^{(k)}$ $(1 \leq k \leq q)$ of the cell, and it is represented as follows:

$$Score(P_j) = \frac{\sum_{k=1}^{q} Var(c^{(k)}, P_j)}{q} \qquad (9)$$

Note that the protein which we want to extract in this study is the protein that controlled score is low.

IV. Evaluation

A. Model of Artificial Data Used in Evaluation

Because no protein controlled by the lineage of Wagyu is known, and so currently there is no real data that we can use to evaluate the proposed algorithm, we generate an artificial data set of proteins and the Wagyu lineages to evaluate the proposed algorithm. In generating an artificial data set, it is significantly important to construct a proper data model that reflects on the real property of the real phenomenon. Thus, we first propose a model of relation between genetic factors in lineage and expression levels of proteins.

In our model, we assume genetic factors that increase/decrease the expression levels of a protein, and they are inherited from ancestors to descendants in the genetic fashion. In general, currently, quantitative traits of creatures as well as protein expressions are regarded to be controlled by plural genetic factors (i.e., genetic polymorphism) [8]. Our assumption is based on this general agreement in the current state of the art.

First, we construct a realistic model of the lineage of Wagyu that properly explains the relation between beef cattle and sires. As described in Section II-A, several excellent sire lines exist in Wagyu as a result of long-time efforts of inbreeding to preserve and improve good genes that generate good quality beef. Thus, we model the sire lines as *sire-trees* that is rooted by a sire and its 10 generation ancestors are included in the tree, as illustrated in Figure 4. We regard that all the sires included in a sire-tree is distinct from others, and we prepare several sire-trees to express several major lines of sires.

Second, we generate and assign genetic factors to the sires in the sire-trees. In this study, we assume that the genetic factors that control the expression levels of a protein are not owned by a small portion of sires, rather broadly owned by sires with a certain probability, although the difference of distribution (i.e., sparse or dense) according to sire lines may be seen.

We designed the genetic factor model as follows: For each protein P_j, we generate r positive genetic factors $g_{j1}^{p}, g_{j2}^{p}, \ldots, g_{jr}^{p}$ and r negative genetic factors $g_{j1}^{n}, g_{j2}^{n}, \ldots, g_{jr}^{n}$, where these positive (resp. negative) genetic factors work to increase (resp. decrease) the expression levels of P_j. Here, note that, genetic factors are generally considered in pairwise fashion due to the pairwise nature of genomes. If we let 'A' be a genetic factor that works to increase/decrease expression levels, and let 'a' be a pairwise component that does not work on expression levels. Then, the genotype can be one of the three types 'AA,' 'Aa,' and 'aa.' We assign each of these genetic factors to every highest generation sire and cows with the genotype 'Aa,' and they are inherited to their descendants according to the law of genetic inheritance. Note that the genetic factors are inherited probabilistically to all the sires in the sire-tree, and the number of genetic factors on each sires is moderately distributed, which includes natural bias that coming from probability nature of inheritance.

Third, we generate beef cattle. The lineage of a beef cattle is uniquely determined if the sires to be ancestor are decided. Now, if we let 1st sire of cattle be its father, let 2nd sire be its mother's father, let 3rd sire be its mother's mother's father, and so on, only we have to do is to determine 1-5th sires for each cattle. So, for each beef cattle, we select 1-5th sires randomly from the sires of 5-10th generations in the generated sire-trees, as illustrated in Figure 5. This operation is done repeatedly for the number of required samples, and then the generated data set is regarded as the lineage data, which is the input of the proposed algorithm.

Finally, we generate protein expression profiles for each cattle. We assume that a protein expression level follow the normal distribution with the average μ and the standard deviation σ. Note that, although protein expression levels are generally regarded to follow log-scale distribution, there is a result in which protein expression levels follow the normal distribution [9]. For the data sets that follow log-scale distribution, we have only to apply logarithm to translate to the normal distribution. The expression level of a protein P_j in a sample i, i.e., $e_{P_j}^{(i)}$, is determined from the base normal distribution and the number of genetic factors corresponding to P_j in a sample (cattle) i. As for the function of genetic factors, we assume that if cattle has 'AA' or 'Aa' genotype as a result of the genetic probabilistic inheritance rule, the genetic factor works to increase/decrease

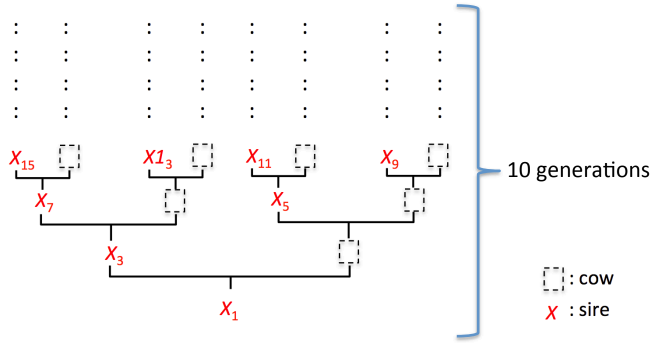

Figure 4. Sire Tree

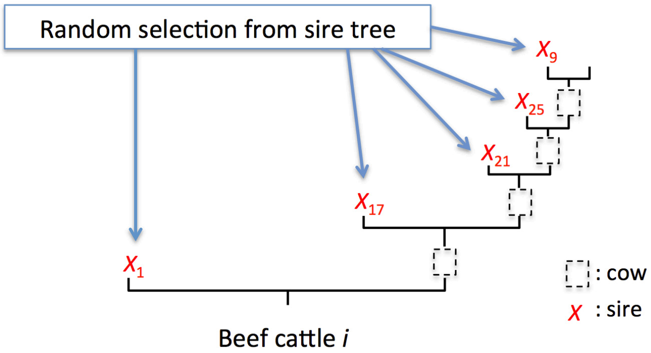

Figure 5. Generating Cattle

expression levels of the corresponding protein. We also assume that, if a genetic factor works, the average of the distribution μ is increased/decreased by a constant amount α. Namely, a set of expression levels corresponding to a protein and a set of samples (cattle) is generated probabilistically based on the normal distribution that average varies according to the number of the genetic factors that the sample have.

As above, we generate the artificial data according to the models of the lineage, the genetic factors, and the protein expression levels. A summary of the generation process of the artificial data is shown in Figure 6.

B. Evaluation Method

We generated a set of artificial data based on the models described in Section IV-A, and applied the proposed algorithm to it. We vary the number of genetic factors corresponding to each proteins; specifically, we choose the number randomly between 0 and 10. Then, if the control scores of a protein computed by the proposed algorithm is in relation to the number of genetic factors corresponding to the protein, it means that the proposed algorithm predicts the number of genetic factors, and further means that the control scores indicate how strong the expression levels of the protein depends on bloodlines.

In the following, we describe the conditions and parameters in the evaluation in detail. Based on the model described in Section IV-A, we generated a set of sire-trees, genetic factors, beef cattle, and expression profiles. In the lineage

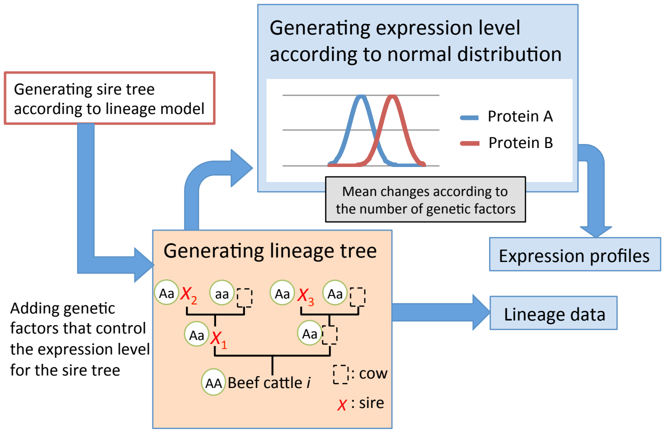
Generating expression level according to normal distribution
Protein A
Protein B
Mean changes according to the number of genetic factors
Generating sire tree according to lineage model
Expression profiles
Generating lineage tree
Aa
X_2
aa
Aa
X_3
Aa
Aa
X_1
Aa
AA
Beef cattle *i*
: cow
X : sire
Adding genetic factors that control the expression level for the sire tree
Lineage data

Figure 6. Process of Generating Artificial Data Set

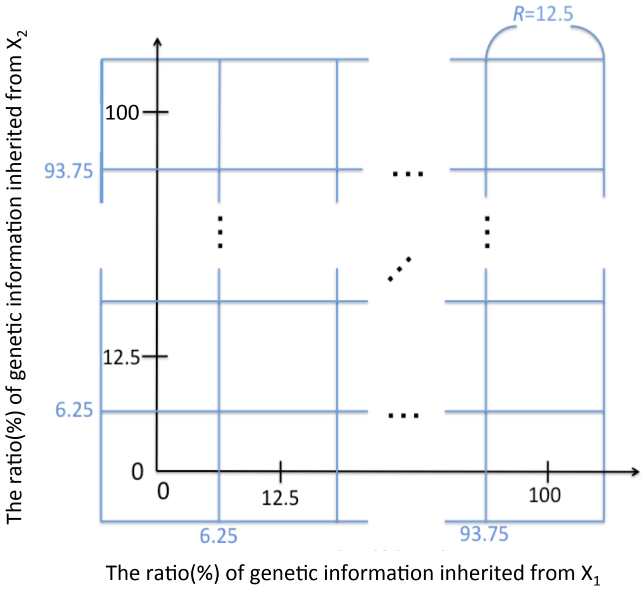
R=12.5
The ratio(%) of genetic information inherited from X_2
100
93.75
12.5
6.25
0
0
12.5
100
6.25
93.75
The ratio(%) of genetic information inherited from X_1

Figure 7. Dividing a Field into Cells

data, we have two sire-trees in which 10 generations of sires are included. The generated set of expression profiles includes 10,000 samples and 100 proteins, where the average and the standard deviation of the base normal distribution are $\mu = 0.5$ and $\sigma = 0.1$, respectively, and we let $\alpha = 1/4\sigma$ be the increment value of the expression level par genetic factor.

The cell width is set to $R = \frac{1}{2^3}$ and the centers of the cells are shifted so that the origin of the field (i.e., the point (0, 0)) is also the center of a cell, as shown in Figure 7. This is meant to have two cattle that has the same 1st, 2nd, and 3rd ancestors is likely to belong to the same cell in many cases. As for the parameters of the algorithm, we set the bandwidth of the Kernel function as $h = 0.21$ to cover cattle in a cell, and set the threshold of the density as $T = 0.7$ in consideration of the distribution of the density with the applied data set.

C. Results

The result of the evaluation is shown in Figure 8. Figure 8 is the scatter diagram where the horizontal axis represents the number of genetic factors and the vertical axis represents the control score, and the plotted points are the proteins. This result shows the strong correlation coefficient -0.838, which means that the proposed method succeeded to estimate the proteins that is controlled by the bloodline.

D. Discussion

By the simulation using an artificial data set, we demonstrated that the proposed method can predict proteins that are deeply related to bloodline. In this simulation, we assumed that each protein has the genetic factors in the DNA of sires, which control the expression level of the protein. This assumption would be widely acceptable because the genetic factors such as SNPs that control phenotypes or expression levels have been explored with tremendous efforts in the current scenes of biological studies.

The result of the simulation showed that the proposed method will work effectively to decide the target proteins to explore the system of living creatures; the protein that has many corresponding genetic factors would be in a position near genetic factors in the biological system, so that the direct interaction between genes and the protein will be found more likely than other proteins. As another practical usage of the proposed method, we suggest the possibility that the proposed method enables us to control important phenotypes more precisely and certainly by selecting better sires for a new-born cattle based on the knowledge of proteins. The proposed method would find proteins controllable by selecting sires, and the proteins that control an important protein will be found by other studies in the future.

There are several possibilities on how to use the knowledge obtained by the proposed method. To explore the practical use of the proposed method is an important task for the future.

V. Conclusion and Future Work

In this paper, we proposed a new method to predict the proteins whose expression levels depend on bloodline, from the lineage data and the protein expression profiles. Because bloodlines include dense genetic information, to connect the bloodline to proteins is valuable when we try to improve the quality of livestock animals through inbreeding. To the best of our knowledge, the proposed method is the first one that investigates the bloodline of brand cattle to connect to proteins.

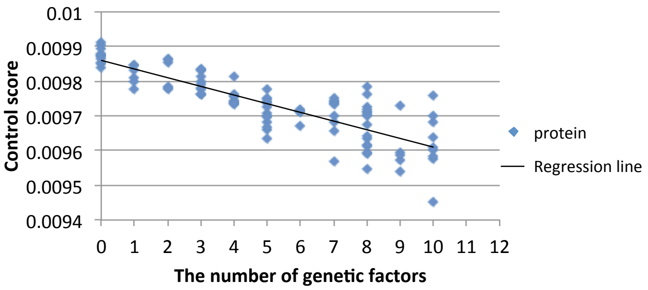

Figure 8. Correlation Between Control Score and Number of Genetic Factors

To evaluate the proposed method, we designed a realistic data model of lineage, genetic factors, and expression profiles, and generated an artificial data. Through the evaluation using the artificial data, we confirmed that the proposed method can find the proteins whose expression levels are controlled by bloodline.

As future work to evaluate the effectiveness of this method firmly, it is desirable to have an evaluation using a real data set. However, the data set that includes an expression profile of proteins and the corresponding bloodline data is not currently available in public. Besides, even if such a data set is available, it is not possible by nature to grasp all the genetic factors that certainly effect on the expression level of a protein. Consequently, it is difficult to evaluate the accuracy of the proposed method exactly.

Considering this difficulty of real-data evaluation, one possible solution would be to demonstrate the effectiveness of the proposed method with some practical case studies. For example, to introduce the case in which the results of the proposed method contributed to a significant discovery or a practical methodology design, would contribute to prove the effectiveness of the proposed method. Although several difficulties are expected, to accumulate an achievement where this methodology worked effectively would be an important task for the future.

Acknowledgement

This work was partly supported by ”the Program for Promotion of Basic and Applied Researches for Innovations in Bio-oriented Industry” of NARO (National Agriculture and Food Research Organization), and ”the Program for Promotion of Stockbreeding” of JRA (Japan Racing Association).

References

[1] N. D. Cameron, “Selection Indices and Prediction of Genetic Merit in Animal Breeding,” CAB International, 1997.

[2] A. M. Campbell and L. J. Heyer, “Discovering Genomics, Proteomics and Bioinformatics,” Benjamin Cummings, 2006.

[3] N. Friedman, M. Linial, I. Nachman, and D. Pe’er, “Using Bayesian Networks to Analyze Expression Data,” Journal of Computational Biology **7**(3/4), pp. 601–620, 2000.

[4] E. Inoue, S. Murakami, T. Fujiki, T. Yoshihiro, A. Takemoto, H. Ikegami, K. Matsumoto, and M. Nakagawa, “Predicting Three-way Interactions of Proteins from Expression Profiles Based on Correlation Coefficient,” IPSJ Transactions on Bioinformatics, Vol. 5, pp. 34–43, 2012.

[5] T. Fujiki, E. Inoue, T. Yoshihiro, and M. Nakagawa, “Prediction of Combinatorial Protein-Protein Interaction from Expression Data Based on Conditional Probability,” Protein-Protein Interactions - Computational and Experimental Tools, InTech Web Press, pp. 131–146, 2012.

[6] C. E. Rasmussen and C. K. I. Williams, “Gaussian Processes for Machine Learning,” the MIT Press, 2006.

[7] R. O. Duda and P. E. Hart, “Pattern Classification and Scene Analysis,” John Wiley & Sons, Inc., 1973.

[8] T. A. Brown, “Genomes 2nd edition,” Garland Science Press, 2002.

[9] N. Balakrishnan and V. B. Nevzorov, “A Primer on Statistical Distributions,” John Wiley & Sons, Inc., 2004.

ISBN: 978-1-61208-295-0