Evolutionary dynamics of spliceosomal intron revealed by *in silico* analyses of the P-Type ATPase superfamily genes

Toshiyuki Oda[1], Ryosuke L. Ohniwa[2], Yuki Suzuki[1], Masatsugu Denawa[1], Masahiro Kumeta[1], Hideyuki Okamura[3] and Kunio Takeyasu[1]

[1]Laboratory of Plasma Membrane and Nuclear Signaling, Kyoto University Graduate School of Biostudies, Yoshida-Konoe-cho, Sakyo-ku, Kyoto 606-8501, Japan

[2]Institute of Basic Medical Sciences, Graduate School of Comprehensive Human Sciences, University of Tsukuba, Tennoh-dai, Tsukuba, Ibaragi 305-8575, Japan

[3]Department of Biology, Osaka Dental University, Kuzuhahanazono-cho 8-1, Hirakata, Osaka 573-1121, Japan

Corresponding author

Toshiyuki Oda

Laboratory of Plasma Membrane and Nuclear Signaling, Graduate School of Biostudies, Kyoto University, Yoshida-Konoe-cho, Sakyo-ku, Kyoto 606-8501, Japan

Phone & Fax: +81-75-753-7905

E-mail: toda.m08@lif.kyoto-u.ac.jp

Abstract

It has been long debated whether spliceosomal introns originated in the common ancestor of eukaryotes and prokaryotes. In this study, we tested the possibility that extant introns were inherited from the common ancestor of eukaryotes and prokaryotes using *in silico* simulation. We first identified 21 intron positions that are shared among different families of the P-Type ATPase superfamily, some of which are known to have diverged before the separation of prokaryotes and eukaryotes. Theoretical estimates of the expected number of intron positions shared by different genes suggest that the introns at those 21 positions were inserted independently. There seems to be no intron that arose from before the diversification of the P-Type ATPase superfamily. Namely, the present introns were inserted after the separation of eukaryotes and prokaryotes.

Keywords

Spliceosomal intron · Introns-late theory · Introns-early theory · Simulation · Proto-splice site · P-Type ATPase

Introduction

Eukaryotic genes often contain non-coding regions called introns that interrupt the protein-coding regions of these genes. Introns are transcribed along with the coding regions into nuclear pre-mRNA and spliced out by the spliceosome, a eukaryota-specific component, during mRNA maturation. These introns are called spliceosomal introns.

Spliceosomal introns are often found at the same positions in homologous genes and even in ancient genes that diverged before the separation of prokaryotes and eukaryotes [1,2]. Therefore, some researchers believe that these introns existed in the common ancestor to prokaryotes and eukaryotes before the divergence. On the other hand, it is known that insertion and deletion of spliceosomal introns occurs frequently [3,4], and, therefore, it is possible that these introns were inserted independently at identical positions in ancient homologous genes after the separation of prokaryotes and eukaryotes [5]. The origin of these introns has long been controversial and is referred to as the "introns-late" versus "introns-early" debate [6].

Several researchers have independently attempted, by statistical approaches, to identify the origin of introns in genes that diverged before and after the separation of prokaryotes and eukaryotes, but their conclusions are inconsistent [1,5,7,8]. These contradictory results may be due to the researchers' distinct assumptions of the positions that allow intron insertion in genes. Iwabe et al. and Kersanach et al. assumed all the nucleotide-nucleotide junctions allow intron insertion [1,8], and their results supported "introns-early" theory. In contrast, "introns-late" theory was supported by Sverdlov et al. , who estimated only "proto-splice sites" are available for intron insertion [7]. Thus, the researchers who set relatively many positions for intron insertion in genes reached "introns-early" theory, and the researchers who set a few positions proposed "introns-late" theory.

In this study, we focused on the P-Type ATPase superfamily, one of the largest gene superfamilies found in both prokaryotes and eukaryotes. This superfamily consists of 5 families. Family 1 comprise the K^+ and heavy metal ion transporters, Family 2 the Ca^{2+} ion transporters, Family 3 the H^+ transporters, Family 4 the lipid translocators, and Family 5 comprise unknown substrates. All the families are found in eukaryotes, and Families 1-3 are also found in prokaryotes [9,10]. Therefore, Families 1-3 are believed to have diverged before the separation of prokaryotes and eukaryotes. Furthermore, although the amino acid sequences share only ~20% identity, the protein structures of all P-Type ATPases are basically similar. Nevertheless, eukaryotic P-Type ATPase genes contain many spliceosomal introns, and some of which exist at the same positions in homologous genes diverged before the separation of prokaryotes and eukaryotes. Now, we can obtain over 100 genes in P-Type ATPase from the gene database. Taking advantage of this large superfamily, we addressed the question about "introns-early" vs "introns-late" theory by our newly established method which enables to eliminate the

unknown variables such as number of positions in the estimation of introns insertion.

Materials and Methods

P-type ATPase superfamily

Gene sequences of the P-Type ATPase superfamily were obtained by database search in GeneBank [11]. The search was conducted in the genomes of 7 animals, 2 plants, 4 fungi, and 5 protists in Dec. 2007. It covered all the eukaryotic organisms whose complete or draft genomes were available at that time. All the P-Type ATPase genes of those organisms referred as RefSeq [12] were collected, except the genes which had stop codons in its coding regions or lacked extremely large regions because these were probably the pseudo genes.

Intron Mapping

The gene sequences were aligned based on their translated amino acid sequences by MAFFT [13], and the alignment was manually modified. Sequences were reverse-translated into nucleotide sequences, and introns were mapped on the alignment.

Note that we applied two different rules, strict and generous rules, for selecting intron positions. Under the strict rule, intron positions in well-aligned regions without any gaps were selected (see Supplemental Data 1). Under the generous rule, all the positions, except the gap-rich regions where intron positions were hardly comparable, were selected.

Genetic Distance

Genetic distance is the phylogenetic distance within a pair of genes calculated by the PROTDIST program in the PHYLIP package [14]. Jones-Taylor-Thornton matrix was used as the scoring matrix.

Index to evaluate the degree of intron inheritance

If introns are independently inserted into two distinct genes, the probability for their insertion at identical positions is calculated as

$$prob_{AB(i)} = ({}_{I_a}C_i \times {}_{(I_{AB\_all}-I_a)}C_{(I_b-i)}) / {}_{I_{AB\_all}}C_{I_b} \qquad \text{Eq (1)}$$

where $prob_{AB(i)}$ is the probability that intron position i is common to genes A and B, I_a and I_b are the numbers of introns in

genes A and B, respectively, and I_{AB_all} is the number of the nucleotide-nucleotide junctions allowing intron insertion in genes A and B. In this formula, all nucleotide-nucleotide junctions are assumed to allow the intron insertion with the same probability.

The expected value for the number of intron positions shared by genes A and B (E_{AB}) is calculated with the following formula.

$$E_{AB} = \sum_{i=1}^{I_b} \left(1 - \left(1 - \frac{1}{I_{AB\_all}}\right)^{n_{Ai}}\right) \qquad \text{Eq (2)}$$

Here, we introduce an index, *Ratio* (*R*), to evaluate the degree of intron inheritance from their ancestor in a pair of genes. *R* for gene pair A and B is calculated as

$$R_{AB} = O_{AB} / E_{AB} \qquad \text{Eq (3)}$$

where O_{AB} is the actual number of intron positions shared by gene pair A and B.

Tree-based weight

Due to the inclusion of highly homologous sequences such as isoforms and paralogs in the sequence population, it is erroneous to treat all of the sequences equally [15]. In this study, we introduced tree-based weight (Fig. 1).

Weight was assigned based on branch length of phylogenetic trees of each family, using kdpB, P-Type ATPase of *Escherichia coli* (NP_415225) as an outgroup. The trees are made by the NEIGHBOR program in the PHYLIP package [14]. Negative branch lengths were set to zero. This method is the same as the method used in CLUSTAL W [16] with the single modification of excluding a normalization step in order to reduce round-off errors.

The tree-based weight of the gene pair A and B is given by

$$W_{AB} = (V_A \times V_B)^{1/2} \qquad \text{Eq (4)}$$

where V_A and V_B are the tree-based weights of genes A and B, respectively.

Simulation test of the *R* values

One 3522 bp (average length of P-type ATPase genes analyzed in this study) gene with 16 introns (average number of introns present in the P-type ATPase genes) was artificially constructed as an ancestral gene for *in silico* simulation. The genes were repeatedly duplicated *in silico* with the mutation rate per nucleotide at 0.001, and the *R* values were calculated under different conditions considering the rate of gain and loss of introns and the position preference of intron insertion. The mutation rate was arbitrarily selected because this rate itself never affected the convergence of the *R* values (data not shown). The detailed procedures are as following.

i) Effect of the rates of gain and loss of introns

After the duplication of the gene, introns were removed on the rates of intron loss (High = 0.6, Middle = 0.4, Low = 0.3 per intron), and new introns were inserted on the rates of intron gain (High = 0.006, Middle = 0.004, Low = 0.003 per nucleotide-nucleotide junction). Position preference was not considered (all the nucleotide-nucleotide junctions allow the intron insertion with the same probability). These steps were repeated 8 times and 256 genes with different number of introns were obtained. Then, *R* values were calculated in each case.

ii) Effect of position preferences for intron insertion

After the duplication of the gene, introns were removed on the rates of intron loss (0.4 per intron). Then, new introns were inserted into duplicated genes on the rates of intron gain (0.004 per nucleotide-nucleotide junction). In this step, the positions of intron insertion were restricted to 25%, 50% and 100% of the nucleotide-nucleotide junctions. These steps were repeated 8 times and 256 genes with different number of introns were obtained. Then, *R* values were calculated in each case.

Results and Discussions

We collected sequences of 354 eukaryotic genes encoding P-Type ATPases by database search, and identified the positions of all introns in those 354 genes (Table 1). We also investigated introns in prokaryotic P-Type ATPase gene and found no introns. Under the strict rule (using well-aligned region among P-Type ATPases. See Intron Mapping section in Materials and Methods), there were 949 introns mapped into 111 positions and under the generous rule (using whole regions except for gap-rich regions in the alignment), there were 4,120 introns mapped into 631 positions (Supplemental Table 1). The 949 introns under the strict rule are a subset of the 4,120 introns under the generous rule. Among the 111 and 631 intron positions identified, 21 and 67 positions were shared by at least two different families, respectively (Fig. 2, Table 2). The numbers 21 and 67 are larger than those reported in similar analyses [1,5,17].

In the case that introns are independently inserted into two different genes, the expected value for "the number of intron positions shared by two genes" can be estimated by Eq (2). If the number of actual intron positions shared by two genes is larger than this expected value, it is reasonable to judge that one or more of the introns have been inherited from a common ancestor (in this paper, we call such introns as "inherited introns"). Therefore, we introduced an index, *Ratio* (*R*) (Eq (3)), to evaluate whether the introns are likely to have been inherited from a common ancestor or to have been inserted independently. Homologous genes that have diverged recently, such as isoforms, are expected to have high *R* values because almost all the introns will be conserved. In contrast, the *R* values for homologous genes that diverged early in the evolution likely to be small, because many insertion/deletion events are thought to have occurred since their divergence. Most of the introns in distant homologues likely to have been inserted after divergence, and, therefore, the O_{AB} values are expected to be close to the E_{AB} values.

The *R* values of all gene pairs (354x353/2=62,481 pairs) were calculated except for 30,096 pairs under the strict rule and 6,688 pairs under the generous rule. To analyze the ancestry of inherited introns, the *R* values were plotted against the genetic distance of each pair, where genetic distance was used to represent the relative time since divergence (Fig. 3a, 3b). The majority of *R* values were almost zero, while some pairs had relatively high *R* values. Since *O* is the actual value in Eq (3) and there will be variation in the *O* value, resulting in scattering in the distribution of *R* values. For example, if genes A and B independently gain 6 and 4 introns, respectively, the possible number of the intron positions shared by the two genes varies from 0 to 4. In this case, there is a high probability of an *R* value of 0 and a very low probability that the *R* value will be extremely high (Table 3). Thus, the relatively high *R* values in Fig. 3a and 3b may not due to the existence of inherited introns, but caused by the existence of introns independently inserted to the identical positions which will occur

with small probability.

Therefore, we averaged the *R* values at every genetic distance to assess the overall trends in *R* dynamics (Fig. 3c, 3d, 3e, 3f). To avoid over-estimation of *R* values associated with the presence of many closely related homologues like isoforms, we applied tree-based weights in the average calculation. The *R* values of intra-family pairs showed a drastic decline with longer genetic distances. This result suggests that recently diverged pairs inherit most of the introns from a common ancestor and that the inherited introns were removed soon after the divergence. In the case of inter-family pairs, the *R* value stays low and constant. This trend did not change when pair selection was restricted to the genes that had diverged before the separation of prokaryotes and eukaryotes (Fig. 3e, 3f). Since the *R* value becomes small when one or both members of pair lose inherited introns, these results indicate that the deletion of inherited introns have occurred quickly under the assumption that new intron insertions/deletions occur continuously, and consequently the *R* values become constant.

According to the definition of *R*, the *R* value should stay as 1 if all the introns are inserted independently. Nevertheless, for most genetic distances, the *R* values of inter-family pairs were greater than 1 (Fig. 3c, 3d, 3e, 3f). There are two possible explanations for such a result. One possibility is, of course, that inherited introns preferentially remain in the genes. In this case, particular pairs of inherited introns should remain regardless of the genetic distance. However, there is no such intron in the P-type ATPases identified (Supplemental Table 1), and, therefore, it is unlikely that introns inherited from an ancestral gene of each family are conserved.

There are several important factors for *in silico* intron analysis such as the rate of gain and loss of introns and the position preference of intron insertions [17, 18]. Here, we simulated how the rate of gain and loss affects to the *R* values with the set of hypothetical genes and introns (Figure 4a). While the rate affects the declining speed of the *R* values, it never affects the convergence values ($R = 1$). Therefore, our observation that the *R* values of P-Type ATPase stayed as over 1 were not be caused by the different rates of gain and loss of the introns.

In contrast, the position preference affected the convergence values of *R* (Fig. 4b). The simulation showed that the *R* values became larger than 1 if the positions of intron insertion were restricted. This is well consistent with our result that *R* value of P-type ATPase stayed at over 1. Therefore, it is likely that there are particular positions that allow intron insertion more frequently than others in P-Type ATPase genes (insertion hot spots).

It has been reported that introns often found at the particular nucleotide sequence: A[A/C]AG![A/G]T [19-21] or !CTC [20] (! is the positions where introns are found). These sequences are believed to act as the “proto-splice site” and allow for easy intron insertion. In the case of the P-Type ATPase, the nucleotide preference at -4 to +4 positions before/after the introns are A[A/C]AG!GTT (Table 4), and are compatible to the reported sequences in other genes [19-21]. In addition, introns are often found in the phase 1 codon of conserved glycines [22]. The P-Type ATPase superfamily has 50 amino acids

conserved in over 80 % members (Table 5), and 11 out of 13 conserved glycine contain introns at Phase 1. Since the glycine codon is GGN, it is reasonable that Phase 1 of glycine naturally constitutes proto-splice sites and potential intron insertion hot spots. Such a relationship between amino acid conservation and intron insertion is also observed with other amino acids, including lysine, which occupies 3 out of the 50 conserved positions in the P-Type ATPase superfamily. The lysine codons are AA[A/G], and phase 3 of all 3 lysine codons have been the target of intron insertion (Table 5). Thus, such conserved amino acids may provide the positions for intron insertion hot spots. Other factors, such as "domains / modules" of proteins [2,23,24] and "phase" in the codons [25] have been reported to affect to the presence of introns. Those factors may also provide the positions for intron insertions hot spots and be responsible for high *R* values.

Concluding Remarks

In conclusion, the introns found at the same positions among the inter-families pairs were not inherited from a common ancestor but were inserted independently after the divergence. Thus, the introns found in the genes belonging to the P-Type ATPase superfamily were independently acquired since the separation of prokaryotes and eukaryotes, supporting the "introns-late" theory. As shown in this study, particular sites on genes accumulate more introns than others. This bias may result in independently inserted introns appearing to be conserved since the separation of prokaryotes and eukaryotes.

References

[1] Kersanach R, Brinkmann H, Liaud M, Zhang D, Martin W, Cerff R (1994) Five identical intron positions in ancient duplicated genes of eubacterial origin. Nature 367:387-389

[2] de Roos A (2007) Conserved intron positions in ancient protein modules. Biol Direct 2:7

[3] Rogozin I, Wolf Y, Sorokin A, Mirkin B, Koonin E (2003) Remarkable interkingdom conservation of intron positions and massive, lineage-specific intron loss and gain in eukaryotic evolution. Curr Biol 13:1512-1517

[4] Nguyen H, Yoshihama M, Kenmochi N (2005) New maximum likelihood estimators for eukaryotic intron evolution. PLoS Comput Biol 1:e79

[5] Rzhetsky A, Ayala F, Hsu L, Chang C, Yoshida A (1997) Exon/intron structure of aldehyde dehydrogenase genes supports the "introns-late" theory. Proc Natl Acad Sci U S A 94:6820-6825

[6] de Souza S (2003) The emergence of a synthetic theory of intron evolution. Genetica 118:117-121

[7] Sverdlov A, Csuros M, Rogozin I, Koonin E (2007) A glimpse of a putative pre-intron phase of eukaryotic evolution. Trends Genet 23:105-108

[8] Iwabe N, Kuma K, Kishino H, Hasegawa M, Miyata T (1990) Compartmentalized isozyme genes and the origin of introns. J Mol Evol 31:205-210

[9] Axelsen K, Palmgren M (1998) Evolution of substrate specificities in the P-type ATPase superfamily. J Mol Evol 46:84-101

[10] Okamura H, Denawa M, Ohniwa R, Takeyasu K (2003) P-type ATPase superfamily: evidence for critical roles for kingdom evolution. Ann N Y Acad Sci 986:219-223

[11] Benson D, Karsch-Mizrachi I, Lipman D, Ostell J, Wheeler D (2007) GenBank. Nucleic Acids Res 35:D21-25

[12] Pruitt K, Tatusova T, and Maglott D (2007) NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res 35: D61-5.

[13] Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30:3059-3066

[14] Felsenstein J (1989) PHYLIP - Phylogeny Inference Package (Version 3.2). Cladistics 5, 164-166.

[15] Vingron M, Sibbald P (1993) Weighting in sequence space: a comparison of methods in terms of generalized sequences. Proc Natl Acad Sci U S A 90:8777-8781

[16] Higgins D, Thompson J, Gibson T (1996) Using CLUSTAL for multiple sequence alignments. Methods Enzymol 266:383-402

[17] Cho G, Doolittle R (1997) Intron distribution in ancient paralogs supports random insertion and not random loss. J Mol Evol 44:573-584

[18] Roy S, Gilbert W (2005) Rates of intron loss and gain: implications for early eukaryotic evolution. Proc Natl Acad Sci U S A 102: 5773-8

[19] Dibb N, Newman A (1989) Evidence that introns arose at proto-splice sites. EMBO J 8:2015-2021

[20] Tomita M, Shimizu N, Brutlag D (1996) Introns and reading frames: correlation between splicing sites and their codon positions. Mol Biol Evol 13:1219-1223

[21] Long M, de Souza S, Rosenberg C, Gilbert W (1998) Relationship between "proto-splice sites" and intron phases: evidence from dicodon analysis. Proc Natl Acad Sci U S A 95:219-223

[22] Endo T, Fedorov A, de Souza S, Gilbert W (2002) Do introns favor or avoid regions of amino acid conservation? Mol Biol Evol 19:521-252

[23] Roy S, Nosaka M, de Souza S, Gilbert W (1999) Centripetal modules and ancient introns. Gene 238:85-91

[24] Kaessmann H, Zollner S, Nekrutenko A, Li W (2002) Signatures of domain shuffling in the human genome. Genome Res 12:1642-1650

[25] Long M, Deutsch M (1999) Association of intron phases with conservation at splice site sequences and evolution of spliceosomal introns. Mol Biol Evol 16:1528-1534

Fig. 1. Scheme of tree-based weights
g1, g2 and g3 represent genes, and b1, b2, b3 and b4 represent provisional branch names. Values in parentheses indicate the branch lengths. The tree-based weight is the summation of the weight of each branch, which is calculated as (length of the branch)/(number of genes which belong to the branch).
For the representative tree shown here, the tree-based weight of g3 is estimated as (length of b2)/2 + (length of b4) = 1.5/2 + 0.5 = 1.25; likewise, that of g1 is 2.0.

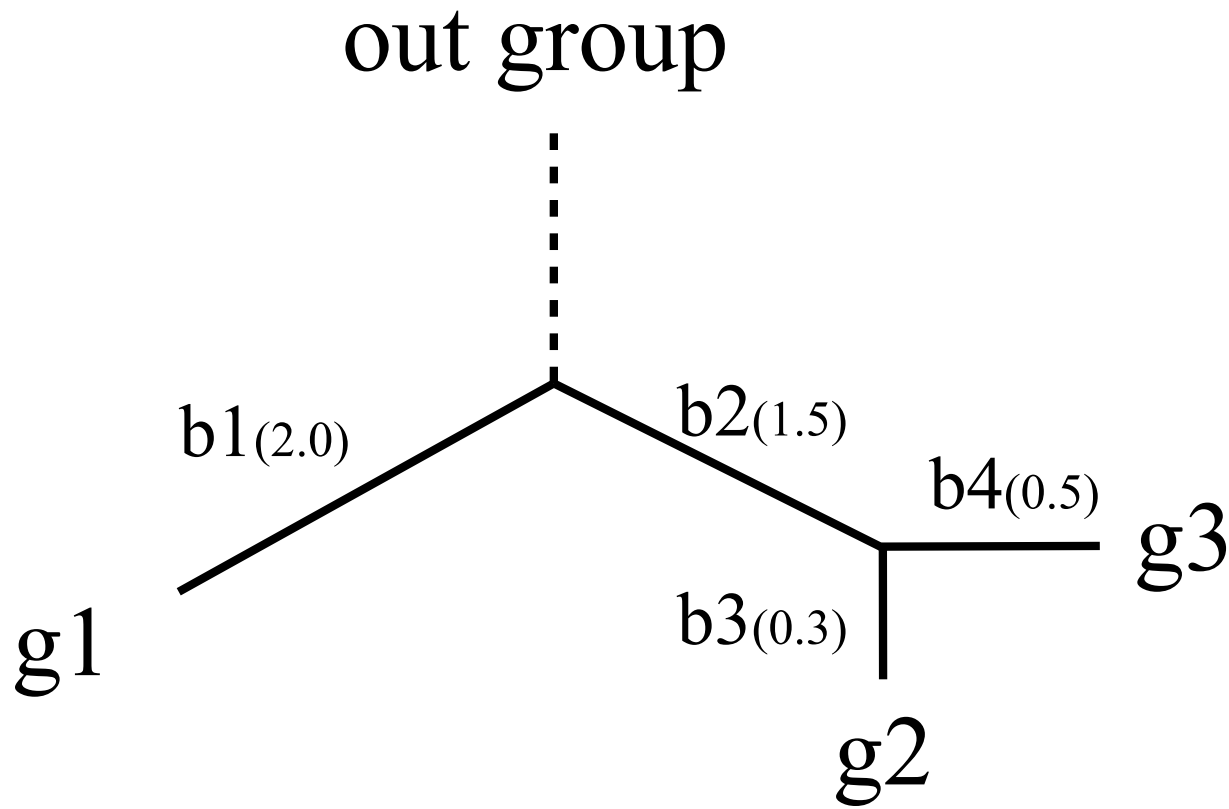
Fig. 2. An example of the intron positions in the alignment of P-Type ATPase.
The first three letters of the gene name represent the species name (osa; *Oryza sativa*, rno; *Rattus norvegicus*, ani; *Aspergillus nidulans*, hsa; *Homo sapiens*, ath; *Arabidopsis thaliana*, dre; *Danio rerio*, gga; *Gallus gallus*, cne; *Cryptococcus neoformans*). Roman numbers of the gene name indicate the family of P-type ATPase to which the genes belong to. (I; Family 1, II; Family 2, III; Family 3, VI; Family 4, V; Family 5). The genes shown here are osa_I (NP_001058417.1), rno_I (NP_036643.1), ani_II (EAA64748.1), hsa_II (NP_001001396.1), ath_III (NP_180028.1), osa_III (NP_001067382.1), ani_IV (EAA58087.1), dre_IV (XP_693773.2), gga_V (NP_001026485.1),and cne_V (XP_570514.1). The BLOSUM 62 score table was used for the shading with GeneDoc package (http://www.nrbsc.org/gfx/genedoc/). The arrows indicate intron positions identified by the strict rule. Colors of the arrows represent the phases of introns; white for phase 0, grey for phase 1 and black for phase 2. Asterisks on the arrows indicate that the introns were conserved across different families.

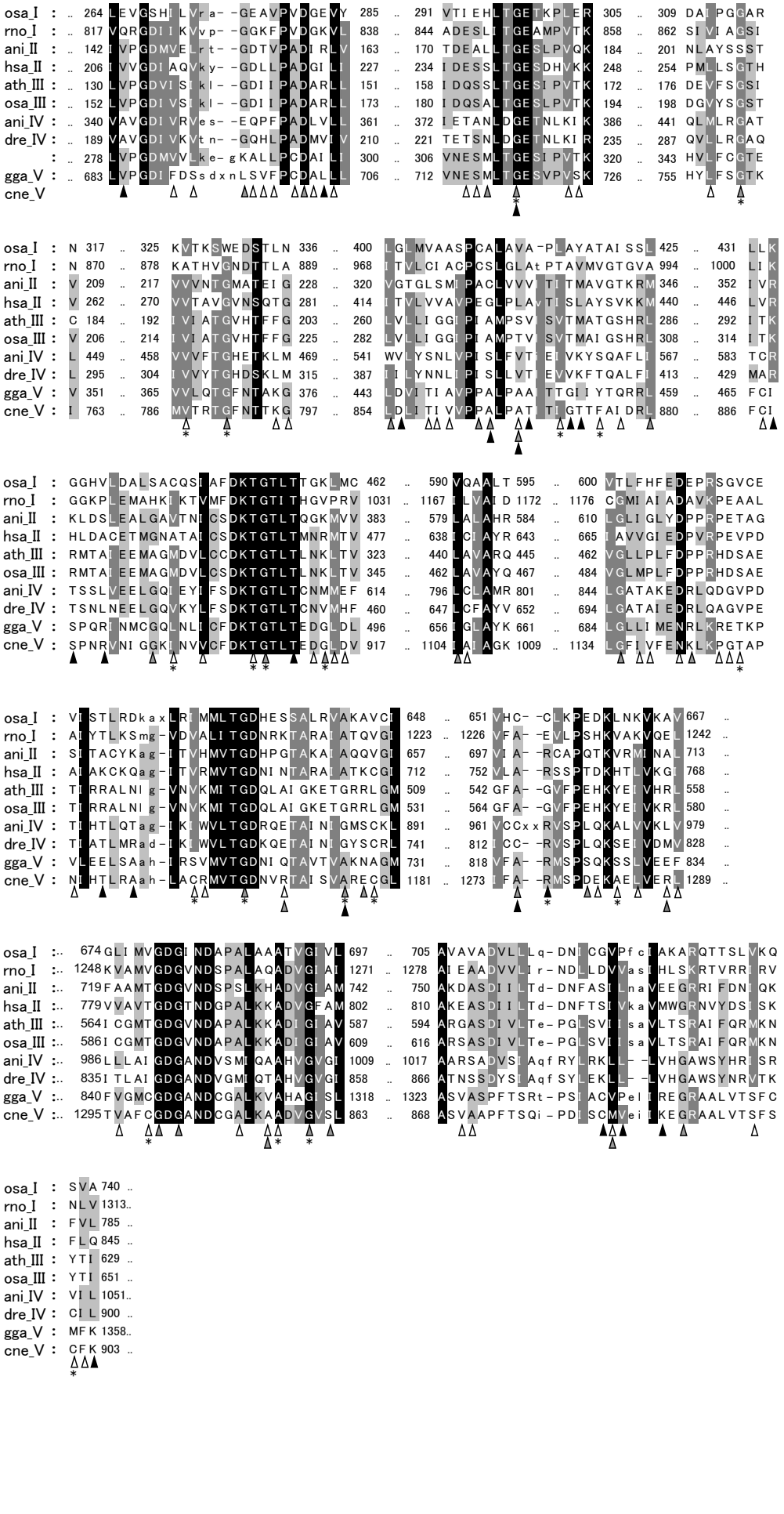
Fig. 3. Correlation between genetic distance and the index, *R*.
a,b; *R* of intra-family pairs (circles) and inter-family pairs (triangles) are plotted against their genetic distances. a represents the plots under the strict rule (intra-family pairs; n=8,895, inter-family pairs; n=23,490). b represents the plots under the generous rule(intra-family pairs; n=16,440, inter-family pairs; n=39,353). c,d; Weighted means of *R* of the intra-family pairs (circles) and the inter-family pairs (triangles), respectively, calculated at genetic distance intervals of 0.2. Error bars indicate ±SE which represents the relative scattering of *R*. Magnifications of the line charts are shown in insets to clarify the trend of the *R* of inter-family pairs. Weighted means of *R* under the strict rule and under the generous rule are shown in c and d, respectively. e, f; *R* of the inter-family pairs were classified into two groups: pairs of genes which had diverged before the separation of prokaryotes and eukaryotes (pairs which members belong to Family 1-2, 2-3 or 1-3, point-down triangles) and the other inter-family pairs (diamonds). e represents *R* under the strict rule (pairs diverged before the separation; n=4,572, other inter-family pairs; n=18,918). f represents *R* under the generous rule (pairs diverged before the separation; n=7,378, other inter-family pairs; n=31,975).

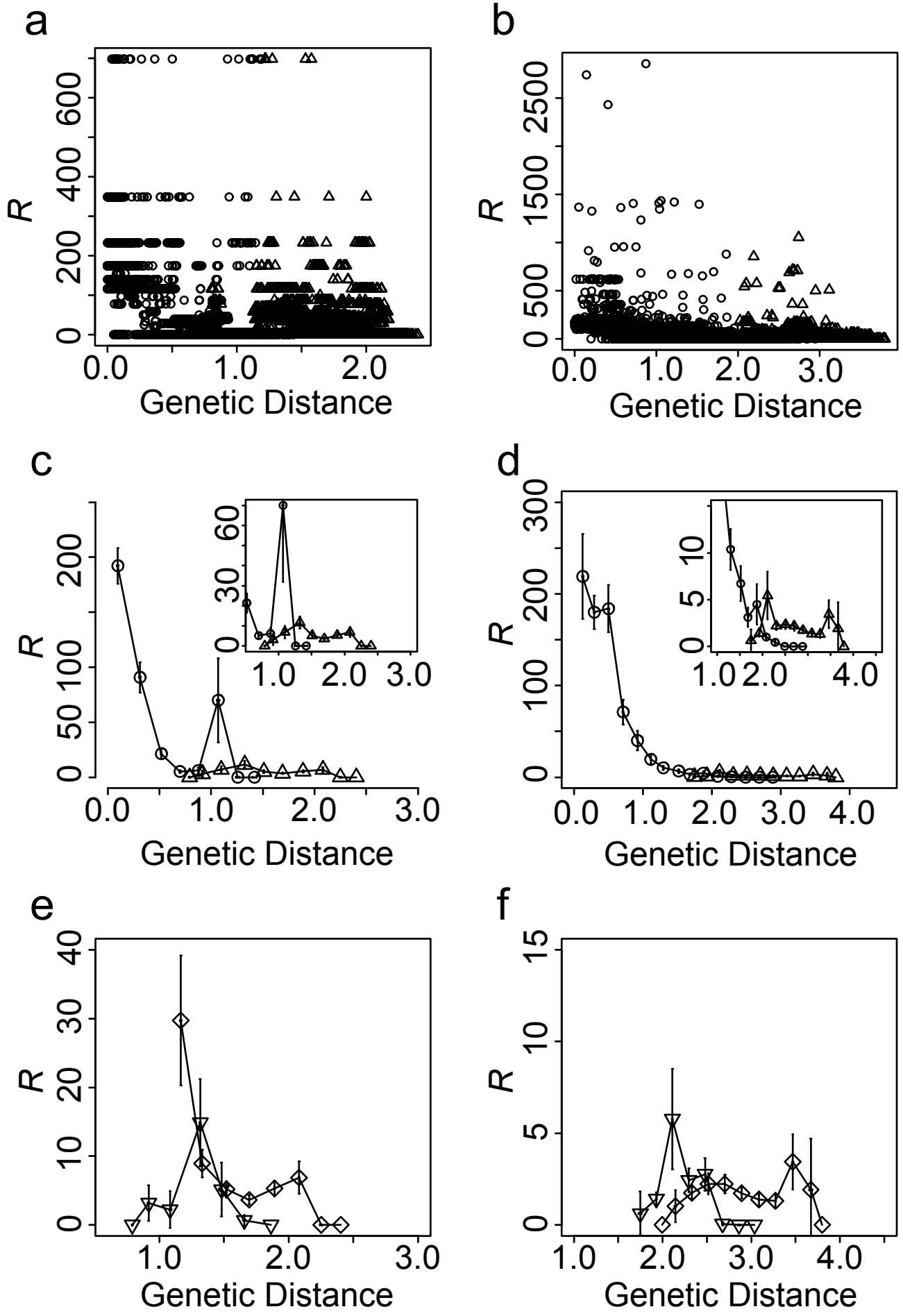
Fig. 4. The effects of “rates of gain and loss of introns" and “position specific preference of intron insertion” on the *R* values.
a; Effects of rates of gain and loss of introns on *R* values. The *R* values of hypothetical genes and introns were simulated under the high (circle and times superimposed), middle (circle and plus superimposed) and low (square and plus superimposed) rates of loss and gain of introns. b; Effects of position specific preference of intron insertions on *R* values. Introns were supposed to be inserted into all the nucleotide-nucleotide junctions (circle and plus superimposed), into 50% of nucleotide-nucleotide junctions (square and times superimposed) and into only 25% of nucleotide-nucleotide junctions (diamonds and plus superimposed).

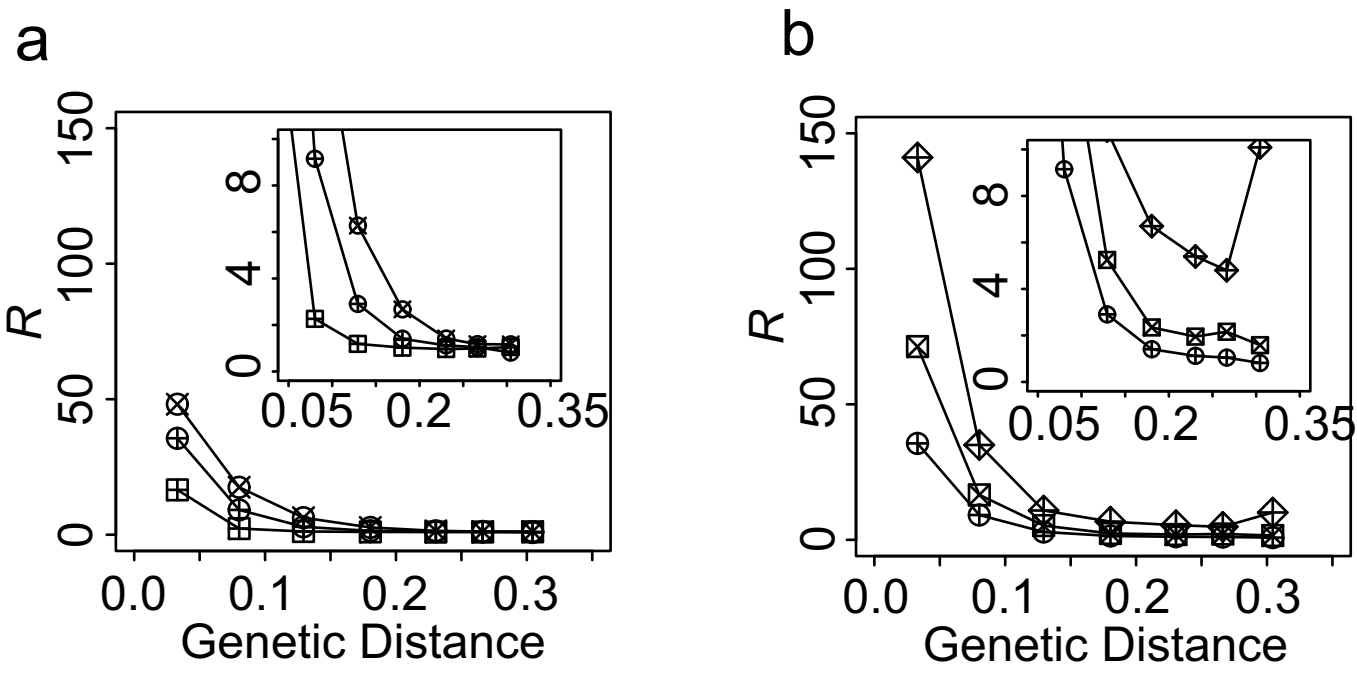
Supplemental Data 1. Alignment of amino acids used in this study.
Amino acids given in upper-case letters were used under the strict rule, while under the generous rule, amino acids given in upper- and lower-case letters, except for “x”, were used. The letter “x” represents amino acids which were not used in any calculations.

Supplemental Table 1. Positions of introns found in the P-Type ATPase genes and the names of genes which have introns at these positions.
The first column indicates positions of nucleotides before the introns, and the second column indicates positions of nucleotides after the introns. The third and subsequent columns indicate the names of genes.


out group
b1(2.0)
b2(1.5)
b4(0.5)
g3
b3(0.3)
g1
g2




a
R
0
200
400
600
0.0
1.0
2.0
Genetic Distance
b
R
0
500
1500
2500
0.0
1.0
2.0
3.0
Genetic Distance
c
R
0
50
100
200
0
30
60
1.0
2.0
3.0
0.0
1.0
2.0
3.0
Genetic Distance
d
R
0
100
200
300
0
5
10
1.0
2.0
4.0
0.0
1.0
2.0
3.0
4.0
Genetic Distance
e
R
0
10
20
30
40
1.0
2.0
3.0
Genetic Distance
f
R
0
5
10
15
1.0
2.0
3.0
4.0
Genetic Distance


a
b
R
Genetic Distance
0
50
100
150
0.0
0.1
0.2
0.3
0
4
8
0.05
0.2
0.35