

**REPORT
OF
THE EXTERNAL REVIEW COMMITTEE
ON
THE MOLECULAR EVOLUTION
RESEARCH PROJECT**



February 2008

**Research Organization of Information and Systems
The Institute of Statistical Mathematics**

【表紙の説明】

正面玄関左壁のレリーフの模写

統計数理の『数』を表し、本研究所の第2代及び第5
代所長 故 末綱 恕一理学博士による揮毫である。

THE INSTITUTE OF STATISTICAL MATHEMATICS

TOKYO
JAPAN

ISM MOLECULAR EVOLUTION RESEARCH PROJECT

REPORT OF THE EXTERNAL REVIEW COMMITTEE

January 2008

FINAL

REPORT

Contents

Forward	1
0. Summary of Recommendations	2
1. The committee and its work	3
1.1 The External Review Committee and its work	3
1.2 Members of the Molecular Evolution Group at ISM	3
1.3 Terms of reference for the committee	4
1.4 Procedures followed by the committee	4
2. Academic Contribution: Introduction and Tribute	5
2.1 Statistical methods: Outcome and impact	6
2.2 Molecular evolution: Outcome and impact	7
2.3 Cooperative Research Activities.	8
3 Contribution to Society	9
4 Bringing up the Next Generation	9
5 Prospects for Future Development and Suggestions	9
5.1 Recommendations	10
Appendix 1: List of Publications from the Molecular Evolution Group at ISM	12
Appendix 2: Schedule of External Review	17

Forward

The Institute of Statistical Mathematics(ISM) was restructured to an independent agency, “Research Organization of Information and Systems”, under the National University Corporation Law in April 2004. Before this restructuring, ISM established a the Prediction and Knowledge Discovery Research Center in 2003 to perform strategically important research area of statistical science. This center studies the statistical modelling and inference algorithms to extract useful information from huge amount of data from complex systems, and aims to solve real world problems in a wide variety of scientific domains, in particular, life, earth, and space sciences. This center is currently composed of the following four groups.

- Molecular Evolution Research Group
- Data Assimilation Research Group
- Statistical Seismology Research Group
- Statistical Genome Diversity Research Group (from 2005)

ISM has carried out assessment of the Molecular Evolution Research Group to review research activities. To this end ISM formed the external review committee consisting of three foreign and one Japanese specialist. The reviewers worked on the related materials which were sent to them beforehand.

The review report was finalized by Prof. David Penny, the chairperson of the committee and the other members of the committee and sent to ISM at the beginning of July 2007. ISM translated this English version into the Japanese version. We must be sorry that publication of the report was too late.

The enormous effort for the review paid by both reviewers and the research group will be rewarded only when the review report results in the improvement of the research system of the molecular evolution research group in ISM. We will consider their evaluation and suggestions seriously and take necessary action to improve and activate the research group.

On behalf of the Institute of Statistical Mathematics, I would like to express my sincere gratitude to the members of the external review committee for their hard and concentrated work.

January 31, 2008

G. Kitagawa
Director-General

Section 0. Summary of Recommendations

The Committee has been extremely impressed by the long-term success of the Molecular Evolution group and thinks that with the continued expansion of genomic data, there will be an increasing requirement for advanced statistical methods in this important area. The current group at the Institute of Statistical Mathematics have been real world leaders in the introduction of full likelihood methods of data analysis into molecular phylogenetics and evolution; they are fulfilling an important role in Japan, and indeed in the world, in this rapidly expanding field. The committee therefore recommends that:

1. a new senior appointment be made to lead the group to work on statistical issues of Molecular Evolution, and the position be advertised internationally,
2. the groups interests be extended to include some more emphasis on statistical computing,
3. the group include programmers whose experience includes graphical interfaces to make software user-friendly to biologists
4. the groups interest be broadened further to include statistical bioinformatics
5. there be an expansion in the number of PhD students.

Section 1. The committee and its work

1.1: The External Review Committee and its work

Membership of Committee (listed alphabetically by family name)

David Penny,	Professor, Massey University, New Zealand (chair).
Andrew J Roger,	Professor, Dalhousie University, Canada
Hidenori Tachida	Professor, Kyushu University, Japan
Ziheng Yang,	Professor, FRS, University College London, United Kingdom.

1.2 Members of the Molecular Evolution Group at ISM

Members at March 2007

Masami Hasegawa	Professor	– Mar 2007
Tetsuo Hashimoto	Associate professor	– Mar 2002
	Visiting professor	Apr 2005 – Mar 2007
Jun Adachi	PostDoc	Jun 1999 – Apr 2000
	Associate professor	Feb 2003 –
Ying Cao	PostDoc	Apr 1998 – Dec 2002
	Assistant professor	Jan 2003 –
Hidenori Nishihara	PostDoc	Apr 2006 – Mar 2007
Yuriko Nishimoto	Graduate student	Dec 2001 – Sep 2003
	PostDoc	Oct 2003 – Mar 2007
Takahiro Yonezawa	Graduate student	Apr 2002 – Mar 2007
Atsushi Matsui	Graduate student	Sep 2004 – Mar 2007

Previous members since the 1998 review

Hidetoshi Shimodaira	Assistant professor	Jul 1996 – May 2002
Peter J. Waddell	PostDoc	Mar 1997 – Mar 1999
Andrew Shedlock	PostDoc	Mar 1998 – Nov 2000
Jeffrey Koji Lum	PostDoc	Mar 1998 – Mar 2000
Margueritte Butler	PostDoc	Mar 1998 – Feb 2000
Christopher Austin	PostDoc	Mar 1998 – Mar 2000
David A. McClellan	PostDoc	Mar 1999 – Feb 2001
Tal Pupko	PosDoc	Nov 2000 – Sep 2002
Masato Nikaido	PostDoc	Apr 2003 – Mar 2006
Rissa Ota	Graduate student	Apr 1998 – Sep 2000
Tae-Kun Seo	Graduate student	Apr 1999 – Sep 2002
Nobuko Arisue	Graduate student	Apr 1999 – Sep 2002
Yumie Murata	Graduate student	Apr 2003 – Mar 2005

1.3 Terms of reference for the committee

The committee was asked to report on the following:

1. Academic contribution in terms of
 - (a) Statistical methods: Outcome and impact
 - (b) Molecular Evolution: Outcome and Impact
 - (c) Cooperative research activities
2. Contribution to Society
3. Bringing up the next generation
4. Prospects for future development and Suggestions.

1.4 Procedures followed by the committee

The committee visited the Institute on 26th & 27th of March 2007. During that time we undertook the following activities and had the following information made available for us.

A detailed report of the past and current activities was prepared by Professor Masami Hasegawa, and a copy of the extensive publication list is attached to this report.

26th March. Symposium: *Towards a New Horizon of Molecular Evolution - Molecular Phylogenetics as a Statistical Science*, held at the Institute on the afternoon of March 26th. The members of the group presented their work during this symposium.

27th March. Members of the external review committee also gave presentations in the morning.

We then met with the Director General, Dr Genshiro Kitagawa, and with Vice-Director General Dr Yoshiyasu Tamura. In the afternoon we had a further opportunity to speak with members of the molecular evolution group, and Professor Kishino from the University of Tokyo also visited the Institute.

In short, the review panel had full access to all the information we required.

Members of the committee would like to express their appreciation to Professor Kitagawa for the opportunity to participate in the Review. We would also like to thank all members of the Institute that we met with, especially to members of the Molecular Evolution group, for the many courtesies extended to us during the visit. We also appreciated the care and thoroughness with which the review materials had been prepared for us before the Review took place.

Section 2. Academic Contribution: Introduction and Tribute

Although our focus is on the period 1998-2007 (the period since the last review) it is definitely appropriate, given the retirement of Professor Masami Hasegawa, to pay a tribute to the contribution of the Molecular Evolution group over the past 25 years. This group have been important world leaders over that time, and probably have been one of the most high-profile and internationally renowned Japanese research groups on any topic. For example, two papers by Hasegawa, Kishino and Yano (1985) and Kishino and Hasegawa (1989) have been cited in the research literature an average of over 2000 times each. This excellent and innovative statistical work has been continued in the past eight years with the publication of some Shimodaira-Hasegawa papers.

In the past twenty years, the Molecular Evolution group has made major contributions to the new and rapidly-expanding field of molecular evolution and phylogenetics. Their contributions have been both in developing statistical methodologies of data analysis, and in applying the methods to sequence data in order to address phylogenetic relationships in several major groups of organisms. As mentioned above, the Markov-process model of nucleotide substitution proposed by scientists in the group in 1984 and 1985, now commonly known as the HKY model, is one of the most widely used models of DNA sequence evolution. The Kishino-Hasegawa test suggested in the group in 1989 is widely used in comparing phylogenetic trees by maximum likelihood. The protein maximum likelihood method of Kishino, Miyata, and Hasegawa has been used frequently for analyzing amino acid sequences, and in particular, in studies of ancient relationships near the root of the tree of life. The MOLPHY program package, developed particularly by Dr Adachi, has found many users and it is excellent to know it is again being updated. It was also Dr Hasegawa's group who introduced the Akaike Information Criterion for model selection to the community of molecular evolutionists and phylogeneticists.

As active researchers ourselves in the field of molecular evolution and systematics, the reviewers have no doubt that research in Dr Hasegawa's group has been at the cutting edge of the field. The group pioneered the use of maximum likelihood methods in the study of molecular phylogeny. In doing this they have made a major contribution to the trend in the study of molecular evolution towards parametric statistical approaches. They were the first major molecular evolution group, both in Japan and in the rest of the world, to take likelihood methods and generalized least squares methods seriously and were able to implement them complex data-structures such as evolutionary trees. As mentioned earlier, they introduced a major stochastic process model of DNA change (the Hasegawa-Kishino-Yano or HKY model) and they explored the use of bootstraps and related tests in studying statistical uncertainty of the phylogenies. This effort has been instrumental in the transformation of the subject area from a subject full of philosophical arguments and *ad hoc* data analysis, to a mature field of applied statistics, with a strong theoretical basis.

The statistical problems that arise involve multiple models (the tree topologies) that are related in highly symmetric but complex ways. The space of tree models is unusual in being a mixture of a discrete and continuous parameters, which fail to meet the usual regularity conditions that are important for most existing statistical methods. At the same time all these probability models fall within the well-investigated exponential family. Thus molecular studies of evolution is a rich source of novel statistical problems, as envisaged by the great statistician Jerzy Neyman (1971). Neyman J. 1971. Molecular studies of evolution: a source of novel statistical problems. Pp. 1-27 in Gupta SS, and Yackel J, eds. Statistical decision theory and related topics. Academic Press, New York.

Besides the many theoretical developments, the group has made crucial contributions to several important biological systems, starting with humans. Examples include phylogenies of humans, apes and the evolution of modern humans, phylogenetic relationships among mammals and vertebrates,

and deep phylogenetic branchings such as the origin of eukaryotes, including a very innovative approach (using duplicated genes) to rooting the tree of life (Iwabe, et al. PNAS 1989). Their effect on biology, beyond methodological questions, has resulted from the energy with which they have addressed a large number of different evolutionary questions with molecular methods. It is arguable that this is the best and most balanced (theoretical and applied) molecular evolution group in the world.

Although the subject area started by applying statistical methods to evolutionary relationships, it is broadening into many other areas of Biology and Medicine. The impact of this group has been on biology and on the development of methods for molecular evolution. Overall, the work of the group has impacted on several areas of biology, from human origins and evolution, to the origin and evolution of mammals, and finally to the deepest divergences of the tree of life, the first divergences after the origin of life itself. However, although the group has been effective in these areas *per se*, their impact has been even more widespread by introducing, and making solid statistical methods available.

A final point in our tribute is that the group has both published in the highest impact journals and has been widely cited in the scientific literature. For example, the group has published in the three top Multidisciplinary journals shown below.

Multidisciplinary Sciences;

Journal	Impact factor	Ranking in field
Science	30.93	1
Nature	29.27	2
Proc. Natl Acad. Sci. USA	10.23	3

Source: *Journal Citation Reports 2005*
Institute for Scientific Information, Philadelphia, PA

Publishing in such journals is highly sought after by researchers.

In addition, they have published in the two highest impact journals for primary research in Evolutionary Biology, namely Systematic Biology (Impact Factor 10.33) and Molecular Biology and Evolution (Impact Factor 6.23). Their work has been highly cited both in Japan and in the rest of the world. To summarize briefly, in terms of

applying fundamental statistics to the novel area of phylogenetics,
in many applications to important evolutionary problems, and
in terms of high impact papers and journals,

the molecular evolution group have been world leaders for over 25 years. We are pleased to pay a tribute to the success of the group. Having done that, we focus more on the eight years since the last external review.

Section 2.1 Statistical methods: Outcome and impact

The Molecular Evolution group has made major contributions in adapting the likelihood method to phylogenetic tree reconstruction, and in particular, in developing statistical tests for evaluating the uncertainties in estimated phylogenies. A phylogeny is best viewed as a model, with its own parameters, so that different phylogenies lie in different parameter spaces. As a result, conventional theory concerning the asymptotic properties of maximum likelihood estimates, does not apply. Evaluating the uncertainties in the estimated model or phylogeny is thus a difficult task. While D.R. Cox discussed some tests comparing different non-nested models, they have never been widely used.

In 1989, Kishino and Hasegawa in the Molecular Evolution group introduced the K-H test, which is based on a normal approximation to the calculated log likelihood difference between two trees.

This test does not accommodate the problem of multiple testing, as may occur when more than two trees are compared. Thus in 1999, Shimodaira and Hasegawa introduced the S-H test, which corrects for multiple testing. Shimodaira (2002) also introduced the almost-unbiased test or AU test. Those papers have by now found widespread use in molecular phylogenetics, attracting thousands of citations. We note that the S-H test and the AU tests were published by the Molecular Evolution group during the period covered by this review. Given the generality of those tests and the current interest in model selection in the statistics literature, we believe those works are important contributions to statistics as well.

Besides tests on trees, the group have also made significant contributions on other statistical problems of molecular evolution and phylogenetics. The group are among the earliest to apply the molecular clock and relaxed molecular clock models to date species divergences, and have continued to make methodological advancements to the problem (Seo et al., 2002; Hasegawa et al. 2003). More recently analysis of large genome-wide datasets is becoming increasingly common. The group evaluated different strategies for analysing genomic data sets (Nishihara et al. 2007), which provide valuable insights for other scientists analyzing large data sets.

Section 2.2 Molecular evolution: Outcome and impact

Here we turn to the application of the methods the group has developed to applications in biology. The first thing to state is that we are impressed how theory and application have supported, and stimulated each other within the group. This leads to application of the powerful statistical methods to real problems that biologists encounter. For example, Hasegawa and Cao, in conjunction with the Okada group at the Tokyo Institute of Technology, are combining sequence data with that from SINES in a novel approach that is increasingly powerful. SINES are 'short interspersed nuclear elements' that can be inserted randomly throughout the genome but can be checked for being unique insertions by sequencing the adjacent DNA region in the different species. Thus it is possible to identify shared insertion events by sequencing the regions adjacent to the insertion site to confirm if the insertion site is the same in all the species. The tree topology for mammals can now also be determined by the SINES, and therefore sequences can then be used independently for estimating the amount (and hence times) of divergence. Simple parsimony is the maximum likelihood estimator for such data (Appl. Math. Lett. 17: 785 2004). We expect to see improved methods that allow SINES and sequence data to be combined formally in a maximum likelihood framework, and that should come soon.

Hasegawa and Hashimoto and their collaborators have also made important contributions to the field of ancient eukaryote phylogeny with the aim to clarify the relationships amongst the simplest nucleus-containing cells. Amongst these manuscripts are a number of important papers describing the red algal phylogeny and the genetic contribution of the chloroplast endosymbiosis on the composition of red algal nuclear genomes (red algae, in the form of 'nori' (*Porphyra yezoensis*) has a special economic importance to Japan). More notable from a pure scientific standpoint is a recent paper by Arisue, Hasegawa and Hashimoto (2005) where the authors used 'separate' maximum likelihood analysis of a large number of genes to attempt to determine the root of the eukaryote tree. This manuscript was particularly noteworthy because of the sophistication of approaches used to identify systematic error in their analyses. It represents one of the only recent analyses to tackling this controversial, challenging and important problem.

In recent years, the molecular evolution group have also made important contributions to the field of viral population genetics and phylogeny. RNA viruses, in particular, evolve very fast and are excellent models for understanding the processes of evolution, and their consequences for disease symptoms. In this viral area they have contributed both novel methodologies in the forms of sampling designs for testing molecular clocks, pseudo-maximum likelihood estimation of ancestral population sizes as well as made important biological discoveries using virus evolution to track human evolution and understand basic molecular biology of viruses. In addition, other areas of

biology have been covered, one of the most important being in using human genetic diversity in populations to follow ancient migration patterns.

In addition to the 'thematic' research topics above, individuals of the group have been involved in diverse 'one-off' collaborative projects that cover an impressive breadth of molecular evolutionary topics. The topics addressed include human diversity studies, cancer biology, parasite evolution and mitochondrial genomics to name a few. Clearly the intellectual environment of the molecular evolution group is stimulating and has fostered a diverse but complementary productive research program. These will be referred to under the next section on Cooperative research activities. Finally, it is excellent to see the group updating some of their earlier computer programs, being able to work at the amino acid level will have several advantages.

Section 2.3 Cooperative Research Activities.

We have already mentioned in the section above that the ISM group has been involved in interactions with other groups. Many of the manuscripts referred to in the foregoing sections are the result of productive collaborations between the statistical molecular evolutionists at ISM and biologists at other institutions in Japan and overseas. Clearly, these other researchers have sought out the unique expertise of the molecular evolution group at the ISM, underlining the group's reputation for excellence.

One of the most direct ways of indicating the extensive set of collaborations with researchers is simply to list different institutions with whom the group has published. For Japan, these include several research institutes (including the National Institute of Basic Biology, Okazaki), Museums, and several other sections of the University of Advanced Study. The universities include the University of Tokyo, Kyoto University, The Tokyo Institute of Technology, Waseda University (Tokyo), Osaka City University, Nagoya University, and Hiroshima University. This extensive list of collaborations within Japan, and some of them have continued over several years, shows again how the group is sought after to assist in both the statistical analysis of DNA data, as well as understanding many of the biological implications. The molecular evolution section of ISM has shown real leadership within Japan.

However, the list is possibly even more extensive outside Japan, even though some collaborations within Japan have been continuing for much longer. Starting within Asia there is the Chinese Academy of Sciences (Wuhan) and the Kyungpook National University in Korea. In Europe the collaborations include Germany, the United Kingdom and Israel. In New Zealand, Otago and Massey Universities, and (among others) Boston University, University of California at Berkeley and the University of Michigan in North America. The conclusion here is the Molecular Evolution group has been doing such excellent and important research that many external researchers wish to collaborate, and that the ISM group has been externally focussed. The molecular evolution group has had an outward-looking focus that has helped get its ideas out to other researchers.

Section 3. Contribution to Society

The most important contribution from scientists to the society is through conducting innovative researches, and breeding the next generation of scientists. In this sense the group has been very successful in contributing to the society as detailed above. Another important aspect of contribution is to make scientists in other fields and general public aware of the scientific results. Although the group has been mostly focusing on innovative researches and generating breeds of young scientists, the group also contributed to the latter aspect.

During this period, Prof. Hasegawa co-authored two Japanese books aimed at scientists in other fields of biology and gave many lectures in other universities as a visiting professor. In addition, the former members of the group are now teaching in major Japanese universities. Finally, public lectures targeted at general public have been given several times by Prof. Hasegawa during the period. Therefore, we think this aspect of contribution also has been made well by this group considering the nature of the institute whose focus is on innovative researches.

Section 4. Bringing up the Next Generation

Here we are particularly interested in both mentoring postdoctoral fellows towards an independent career, and in graduate education. Although our committee was charged primarily with examining the research activities of this group, we think that it is appropriate for us to make some comments on the role of this group in graduate education. We do not see as many Japanese graduate students as we expected taking advantage of the presence of this group. The ISM molecular evolution group is so strong that Japanese graduate students should be encouraged to learn from it, and use it as a basis for their study. Our conclusion was reinforced by a visit by three members of the external review committee to China in early June, where we found a really major push to increase PhD training. The formation of the Graduate University for Advanced Study has been a major advantage but certainly more graduate students in the Institute are highly desirable.

Section 5. Prospects for Future Development and Suggestions

In addition to our comments on the past work of the Molecular Evolution group, it is certainly appropriate to consider future possibilities for ways in which the work of this excellent group can have an even greater effect. The first remark is that it is always necessary to consider new opportunities that result from new developments in statistics, computer science, and especially the huge expansion of genomic data, which is expected to keep increasing. It is this last aspect that is particularly striking. The ‘flood’ of genomic DNA sequences will almost certainly keep expanding as new technologies come on stream. Turning this flood of data into real biological understanding is increasingly going to require improved statistical and computational methods, combined with a positive interaction with biologists.

(1) Increased connection with computational molecular biology. The current group is very strong, as we have seen, in empirical work in molecular evolution and in the application of novel statistical methods to that field. Perhaps not as much effort has been devoted to its connection to the growing field of computational molecular biology (part of bioinformatics). Computer programs for statistical analysis of phylogenies have been developed and distributed by Adachi and Hasegawa (their package MolPhy). But as interest in computational molecular biology has increased rapidly among computer scientists in the past few years, now would be an excellent moment to expand the group's contacts with computer scientists in the Tokyo area. There are a variety of interesting questions -- such as the combined ‘alignment’ and tree inference problem -- these are of great interest and this group could act as a stimulus to Japanese computer scientists to take up these

problems as well. One difficulty computer scientists elsewhere have had is an unwillingness to consider these as statistical problems. The fact that this group is located in a statistics institute may influence Japanese computer scientists to view these problems statistically as well as algorithmically.

(2) Closer connection between the work on molecular evolution and statistical theory.

The presence of strong statisticians like Dr Shimodaira (and in the past, Prof. Kishino) has greatly helped the work of this group. We see the strong statistical component as crucial to the future of the group, and very important for the subject as a whole. Care needs to be taken to see that future statistical methodology continues to be responsive to the needs of the data analysis. A possible scenario to be avoided is to have the work on statistical methods gradually deviate from the needs of the data analysis. One possible way of encouraging co-ordination would be to have statistics graduate students supervised by Dr. Shimodaira work on problems of practical importance to the data analysis in molecular evolution.

Section 5.1. Recommendations

The committee has been very impressed by the work of the Molecular Evolution group over the last 25 years. Given the retirement of Professor Hasegawa, the long-term leader of the group, it is certainly appropriate to offer suggestions for allowing the ISM to continue to be world leaders in this expanding area. The committee therefore makes the following recommendations, and we consider the recommendations to be a tribute to the success of the group.

1. That a new senior appointment be made in fundamental statistical aspects of Molecular Evolution, and the position be advertised internationally.

The committee considered that there is an excellent nucleus for renewing the group, allowing some repositioning as statistical methods become more powerful and the needs of the user community keep adapting to new opportunities (both computing and the large increase in DNA sequences). We see the need for a new leader with a strong theoretical statistical background who will inspire the group as a whole toward new challenges. There certainly are Japanese candidates who may be attracted to the position, but as a matter of principle we think that the position should be advertised internationally. It may be possible to attract an overseas (Japanese) researcher, but even if a appointment is made from within Japan it would give the new group Director the confidence that they were appointed after an international search.

2. That the groups interests be extended to include some more emphasis on statistical computing.

The newer methods of analysis are requiring considerable computing time, and are exploring a wider range of trees and models. In the introduction to this section (section 5) we gave some of the reasoning behind this conclusion. Our datasets are becoming very much larger, and it is reasonable to estimate some additional parameters for the models – this makes it easier to include additional biological realism. Researchers are asking additional questions, including which sites of a gene may have been the target for positive selection. In addition, it is becoming more reasonable to infer additional regulatory signals in the data, and this allows much better estimation of how, for example, the information in a genome will translated into a functioning cell through multiple layers of control.

3. That the group include programmers whose experience includes graphical user interfaces (GUI).

It appeared to the committee members that some talented researchers in the group were having to spend considerable time on the complex tasks of software implementation of GUIs and developing applications that were ‘platform independent’ – that is, usable on Windows, Unix, Linux, and MacOSX operating systems. Such tasks are complex in their own right but are more

appropriate for a dedicated programmer, freeing up the researchers to focus more on methodological innovation as well as evaluation and testing with real data sets. This recommendation is clearly related to number 2, we just do not see that it is possible for a small group of programmers to do everything.

4. That the groups' interest be broadened further to include statistical bioinformatics.

With the advent of more completely sequenced genomes there is an increasing demand for better statistical analysis. Again, this concept was introduced in the introduction to Section 5, and this recommendation follows on from numbers 2 and 3. Biologists can be very innovative in designing their laboratory experiments, but they are simply not equipped to be innovative statisticians to come up with new types of analyses. For certainly the next 20-30 years we see a major requirement for statisticians able to work with biologists in getting much more information out of genomes. It is certainly something that biologists are not going to be able to do on their own. There are major opportunities here that will only keep expanding. To take just one example, we have little idea of some particular mutations lead to disease symptoms – there is an almost unlimited room for expansion in applying innovative statistical thinking to get at some of the vast amount of unused information in genomes.

5. That there be an expansion in the number of PhD students.

The subject of Molecular Evolution is excellent for graduate education in that it is good training for some complex aspects of statistical inference. The ISM seemed to the committee to be an ideal location for graduate education, and emphasizing graduate education may also make the position of a new group Director even more desirable, something that can only help in the recruitment process. Ideally, we see the expansion in both Japanese and non-Japanese students, but an expansion of non-Japanese students may depend on Government policy. There is an international trend for governments to help ensure quality of research institutions by measuring their ability to attract PhD students from outside the country. We can only reiterate that molecular evolution is an excellent topic for the education of graduate students (it keeps the students thinking broadly and quantitatively) and that the Institute of Statistical Mathematics is an excellent location for this process.

Appendix1

List of Publications from the Molecular Evolution Group at ISM 1999 – 2007

The papers with * are included in the collection of publications.

Statistical methods

1. *H. Shimodaira and M. Hasegawa (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol. Biol. Evol.*, 16:1114-1116.
2. *J. Adachi, P. Waddell, W. Martin and M. Hasegawa (2000) Plastid genome phylogeny and a model of amino acid substitution for proteins encoded by chloroplast DNA. *J. Mol. Evol.*, 50:348—358.
3. *R. Ota, P.J. Waddell, M. Hasegawa, H. Shimodaira and H. Kishino (2000) Appropriate likelihood ratio tests and marginal distributions for evolutionary tree models with constraints on parameters. *Mol. Biol. Evol.*, 17:798—803.
4. D. Penny and M. Hasegawa (2001) Covarion model of molecular evolution. In S. Brenner and J.H. Miller, eds., *Encyclopedia of Genetics*, pp. 473—477, Academic Press (San Diego).
5. H. Kishino and M. Hasegawa (2001) Maximum likelihood. In S. Brenner and J.H. Miller, eds., *Encyclopedia of Genetics*, pp. 1157—1160, Academic Press (San Diego).
6. *H. Shimodaira and M. Hasegawa (2001) CONSEL: a program for assessing the confidence of phylogenetic tree selection. *Bioinformatics* 17: 1246-1247.
7. H. Shimodaira (2001) Multiple comparisons of log-likelihoods and combining nonnested models with applications to phylogenetic tree selection. *Comm. Statist. Theory Methods*, 30:1751-1772.
8. D.A. McClellan and K.G. McCracken (2001) Estimating the influence of selection on the variable amino acid sites of the cytochrome *b* protein functional domains. *Mol. Biol. Evol.*, 18:917-925.
9. *H. Shimodaira (2002) An approximately unbiased test of phylogenetic tree selection. *Syst. Biol.*, 51:492-508.
10. *T. Pupko, D. Huchon, Y. Cao, N. Okada, and M. Hasegawa (2002) Combining multiple data sets in a likelihood analysis: Which models are the best? *Mol. Biol. Evol.*, 19:2294-2307.
11. *T. Pupko, I. Pe'er, M. Hasegawa, D. Graur, and N. Friedman (2002) A branch-and-bound algorithm for the inference of ancestral amino-acid sequences when the replacement rate varies among sites: Application to the evolution of five gene families. *Bioinformatics*, 18: 1116-1123
12. *T. Pupko, R. Sharan, M. Hasegawa, R. Shamir, and D. Graur (2003) Detecting excess radical replacements in phylogenetic trees. *Gene* 319: 127-135.
13. P.J. Waddell, H. Mine, and M. Hasegawa (2004) INTEROGATE 1.0. Exploration and testing of stationarity, reversibility and clock-likeness in sequence data. *Computer Science Monograph (ISM)*, 31:1-22.
14. *H. Shimodaira and M. Hasegawa (2005) Assessing the uncertainty in phylogenetic inference. In R. Nielsen, ed., *Statistical Methods in Molecular Evolution*, pp.463—493, Springer.
15. *K. Shimizu, J. Adachi, and Y. Muraoka (2006) ANGLE: A sequencing errors resistant program for predicting protein coding regions in unfinished cDNA. *J. Bioinf. Comput. Biol.*, 4:649-664.
16. Z. Li, Q. Liu, M. Song, Y. Zheng, P. Nan, Y. Cao, G. Chen, Y. Li, and Y. Zhong (2005) Detecting correlation between sequence and expression divergences in a comparative analysis of human serpin genes. *BioSystems* 82:223- 226.

Vertebrate evolution

17. *P. Waddell, N. Okada, and M. Hasegawa (1999) Towards resolving the interordinal relationships of placental mammals. *System. Biol.*, 48:1-5.
18. *P. Waddell, Y. Cao, J. Hauf, and M. Hasegawa (1999) Using novel phylogenetic methods to evaluate mammalian mtDNA, including amino acid-invariant sites-LogDet plus site stripping, to detect internal conflicts in the data, with special reference to the positions of hedgehog, armadillo, and elephant. *System. Biol.*, 48: 31-53.
19. D. Penny, M. Hasegawa, P.J. Waddell, and M.D. Hendy (1999) Mammalian evolution: Timing and implications from using the LogDeterminant transform for proteins of differing amino acid composition. *System. Biol.*, 48:76-93.
20. *P. Waddell, Y. Cao, M. Hasegawa, and D.P. Mindell (1999) Assessing the Cretaceous superordinal divergence times within birds and placental mammals by using whole mitochondrial protein sequences and an

- extended statistical framework. *System. Biol.*, 48:119-137.
21. *D.P. Mindell, M.D. Sorenson, D.E. Dimcheff, M. Hasegawa, J. Ast, and T. Yuri (1999) Interordinal relationships of birds and other reptiles based on whole mitochondrial genomes. *System. Biol.*, 48:138-152.
 22. *Y. Cao, K.S. Kim, J.H. Ha and M. Hasegawa (1999) Model dependence of the phylogenetic inference: Relationship among carnivores, perissodactyls and cetartiodactyls as inferred from mitochondrial genome sequences. *Genes Genet. Syst.*, 74: 211-217.
 23. C.C. Austin (1999) Lizards took express train to Polynesia (1999) *Nature*, 397: 113-114
 24. C.C. Austin (1999) Island colonization by *Lipinia noctua* (Reptilia: Scincidae) in Melanesia: Molecular phylogeny and population structure based on mitochondrial cytochrome b and 12S rRNA Genes. *Island Herpetofauna: Origin, Current Status and Conservation* (ed. T. Ota), Elsevier Press, pp. 169—189.
 25. C.C. Austin (1999) Colonization, isolation and speciation of lizards on Pacific islands: a molecular genetic perspective. *Aquabiology*, 21:313-318.
 26. C.C. Austin and G. Zug (1999) Molecular and morphological evolution in the south-central Pacific skink *Emoia tongana* (Reptilia: Squamata): uniformity and human-mediated dispersal. *Austral. J. Zool.*, 47: 425-437.
 27. Y. Tohyama, H. Kasama-Yoshida, M. Sakuma, Y. Kobayashi, Y. Cao, M. Hasegawa, H. Kojima, Y. Tamai, M. Tanokura, and T. Kurihara (1999) Gene structure and amino acid sequence of Latimeria chalumnae (coelacanth) myelin DM20: phylogenetic relation of the fish. *J. Neurochem.*, 24:867—873.
 28. *J.K. Lum, M. Nikaido, M. Shimamura, H. Shimodaira, A.M. Shedlock, N. Okada, and M. Hasegawa (2000) Consistency of SINE insertion topology and flanking sequence tree: Quantifying relationships among cetartiodactyls. *Mol. Biol. Evol.*, 17:1417-1424.
 29. *M. Nikaido, M. Harada, Y. Cao, M. Hasegawa and N. Okada (2000) Monophyletic origin of the order Chiroptera and its phylogenetic position among Mammalia, as inferred from the complete mitochondrial DNA sequence of a Japanese megabat, the Ryukyu flying fox (*Pteropus dasymallus*). *J. Mol. Evol.*, 51:318-328.
 30. *Y. Cao, M.D. Sorenson, Y. Kumazawa, D.P. Mindell and M. Hasegawa (2000) Phylogenetic position of turtles among amniotes: evidence from mitochondrial and nuclear genes. *Gene*, 259:139-148.
 31. *Y. Cao, M. Fujiwara, M. Nikaido, N. Okada and M. Hasegawa (2000) Interordinal relationships and time-scale of eutherian evolution as inferred from mitochondrial genome data. *Gene*, 259:149-158.
 32. A.M. Shedlock, M.C. Milinkovitch, and N. Okada (2000) SINE evolution, missing data, and the origin of whales. *Syst. Biol.* 49:808-817.
 33. A.M. Shedlock and N. Okada (2000) SINE insertions: powerful tools for molecular systematics. *BioEssays*, 22:148-160.
 34. M.A. Butler, T.W. Schoener, and J.B. Losos (2000) The relationship between habitat type and sexual size dimorphism in Greater Antillean Anolis lizards. *Evolution*, 54:259-272.
 35. Y. Tohyama, T. Ichimiya, H. Kasama-Yoshida, Y. Cao, M. Hasegawa, H. Kojima, Y. Tamai, and T. Kurihara (2000) Phylogenetic relation of lungfish indicated by the amino acid sequence of myelin DM20. *Mol. Brain Res.*, 80:256—259
 36. M. Onuma, Y. Cao, M. Hasegawa and S. Kusakabe (2000) A close relationship of Chiroptera with Eulipotyphla (core Insectivora) suggested by four mitochondrial genes. *Zool. Sci.*, 17:1327—1332.
 37. S. Ohdachi, N.E. Dokuchaev, M. Hasegawa, and R. Matsuda (2001) Intraspecific phylogeny and geographic variation of six species of northern Asiatic *Sorex* shrews based on the mitochondrial cytochrome b sequences. *Mol. Ecol.* 10:2199—2213
 38. M. Sumida, Y. Kanamori, H. Kaneda, Y. Kato, M. Nishioka, M. Hasegawa, and H. Yonekawa (2001) Complete nucleotide sequence and gene rearrangement of the mitochondrial genome of the Japanese pond frog *Rana nigromaculata*. *Gene Genet. Syst.* 76:311—325
 39. *M. Nikaido, F. Matsuno, H. Hamilton, R.L. Brownell Jr., Y. Cao, D. Wang, Z. Zhu, A.M. Shedlock, R.E. Fordyce, M. Hasegawa, and N. Okada (2001) Retroposon analysis of major cetartean lineages: The monophyly of toothed whales and the paraphyly of river dolphins. *Proc. Natl. Acad. Sci. USA*, 98: 7384-7389.
 40. *M. Nikaido, K. Kawai, Y. Cao, M. Harada, S. Tomita, N. Okada, and M. Hasegawa (2001) Maximum likelihood analysis of the complete mitochondrial genomes of eutherians and a reevaluation of the phylogeny of bats and insectivores. *J. Mol. Evol.* 53:508-516
 41. K. Kawai, M. Nikaido, M. Harada, S. Matsumura, L.-K. Lin, Y. Wu, M. Hasegawa, and N. Okada (2002) Intra- and interfamily relationships of Vespertilionidae inferred by various molecular markers including SINE insertion data. *J. Mol. Evol.*, 55: 284-301
 42. M.A. Butler and J.B. Losos (2002) Multivariate sexual dimorphism, sexual selection, and adaptation in Greater

Antillean Anolis lizards. *Ecological Monographs*. 72:541-559.

43. *M. Nikaido, Y. Cao, N. Okada, and M. Hasegawa (2003) The phylogenetic relationships of insectivores with special reference to the lesser hedgehog tenrec as inferred from the complete sequence of their mitochondrial genome. *Genes Genet. Syst.* 78: 107-112.
44. *M. Hasegawa, J.L. Thorne, and H. Kishino (2003) Time scale of eutherian evolution estimated without assuming a constant rate of molecular evolution. *Genes Genet. Syst.* 78: 267-283
45. *Y. Murata, M. Nikaido, T. Sasaki, Y. Cao, Y. Fukumoto, M. Hasegawa, and N. Okada (2003) Afrotherian phylogeny as inferred from complete mitochondrial genomes. *Mol. Phylogenet. Evol.* 28: 253-260.
46. *M. Nikaido, Y. Cao, M. Harada, N. Okada, and M. Hasegawa (2003) Mitochondrial phylogeny of hedgehogs and monophyly of Eulipotyphla. *Mol. Phylogenet. Evol.* 28: 276-284.
47. K. Kawai, M. Nikaido, M. Harada, S. Matsumura, L.-K. Lin, Y. Wu, M. Hasegawa, and N. Okada (2003) The status of the Japanese and East Asian bats of the genus *Myotis* (Vespertilionidae) based on mitochondrial sequences. *Mol. Phylogenet. Evol.*, 28: 297-307.
48. A.M. Shedlock, T.W. Pietsch, M.G. Haygood, P. Bentzen, M. Hasegawa (2004) Molecular systematics and life history evolution of anglerfishes (Teleostei: Lophiiformes): Evidence from mitochondrial DNA. *Steenstrupia*, 28:129-144.
49. *T. Sasaki, M. Nikaido, H. Hamilton, M. Goto, H. Kato, N. Kanda, L.A. Pastene, Y. Cao, R.E. Fordyce, M. Hasegawa, and N. Okada (2005) Mitochondrial phylogenetics and evolution of Mysticete whales. *Syst. Biol.*, 54: 77-90.
50. Butler M.A. (2005) Foraging mode of the chameleon, *Bradypodion pumilum*: a challenge to the sit-and-wait versus active forager paradigm? *Biol. J. Linnean Soc.*, 84:797-808
51. *A. Kurabayashi, C. Usuki, N. Mikami, T. Fujii, H. Yonekawa, M. Sumida, and M. Hasegawa (2006) Complete nucleotide sequence of the mitochondrial genome of a Malagasy poison frog *Mantella madagascariensis*: Evolutionary implications on mitochondrial genomes of higher anuran groups. *Mol. Phylogenet. Evol.* 39: 223-236
52. *H. Nishihara, M. Hasegawa, and N. Okada (2006) Pegasoferae, an unexpected mammalian clade revealed by tracking ancient retroposon insertions. *Proc. Natl. Acad. Sci. USA* 103: 9929-9934
53. H. Endo, T. Yonezawa, F. Rakotondraparany, M. Sasaki, and M. Hasegawa (2006) The adaptational strategies of the hindlimb muscles in the Tenrecidae species including the aquatic web-footed tenrec (*Limnogale mergulus*). *Annals of Anatomy* 188:383-390
54. S.D. Ohdachi, M. Hasegawa, M.A. Iwasa, P. Vogel, T. Oshida, L.-K. Lin and H. Abe (2006) Molecular phylogenetics of soricid shrews (Mammalia) based on mitochondrial cytochrome *b* gene sequences: with special reference to the Soricinae. *J. Zool.*, 270:177-191 (Errata 270:199-200).
55. *T. Sasaki, M. Nikaido, S. Wada, T.K. Yamada, Y. Cao, M. Hasegawa, and N. Okada (2006) *Balaenoptera omurai* is a newly discovered baleen whale that represents an ancient evolutionary lineage. *Mol. Phylogenet. Evol.* 41:40-52
56. *M. Watanabe, M. Nikaido, T.T. Tsuda, T. Kobayashi, D. Mindell, Y. Cao, N. Okada, and M. Hasegawa (2006) New candidate species most closely related to penguins. *Gene* 378: 65-73.
57. *M. Munemasa, M. Nikaido, S. Donnellan, C.C. Austin, N. Okada, and M. Hasegawa (2006) Phylogenetic analysis of diprotodontian marsupials based on complete mitochondrial genomes. *Genes Genet. Syst.* 81:181-191.
58. L.A. Pastene, M. Goto, A.N. Zerbini, K. Watanabe, Y. Bessho, D. Kerem, N. Kanda, M. Hasegawa, R. Nielsen, F. Larsen, and P.J. Palsbøll (2007) Marine speciation during periods of global warming: timing of origin and radiation of the common minke whale, *Balaenoptera acutorostrata*. *Molecular Ecology* (in press)
59. *A. Matsui, F. Rakotondraparany, M. Hasegawa, and S. Horai (2007) Determination of a complete lemur mitochondrial genome from feces. *Mammal Study* (in press)
60. *T. Yonezawa, M. Nikaido, N. Kohno, Y. Fukumoto, N. Okada, and M. Hasegawa (2007) Molecular phylogenetic study on the origin and evolution of Mustelidae. *Gene* (in press)
61. *H. Nishihara, N. Okada, and M. Hasegawa (2007) Rooting the eutherian tree—The power and pitfalls of phylogenomics (submitted).

Eukaryote evolution

62. M. Hasegawa and T. Hashimoto (1999) Phylogenetic position of amitochondriate protists in the evolution of eukaryotes. *Biol. Bull.*, 196: 389—392.

63. *N. Arisue, T. Hashimoto, J.A. Lee, D.V. Moore, P. Gordon, C.W. Sense, T. Gaasterland, M. Hasegawa, and M. Muller (2002) The phylogentic position of the pelobiont *Mastigamoeba balamuthi* based on sequences of rDNA and translation elongation factors EF-1 α and EF-2. *J. Euk. Microbiol.*, 49:1-10.
64. *N. Arisue, T. Hashimoto, H. Yoshikawa, Y. Nakamura, G. Nakamura, F. Nakamura, T. Yano, and M. Hasegawa (2002) Phylogenetic position of *Blastocystis hominis* and of stramenopiles inferred from multiple molecular sequence data. *J. Eukaryot. Microbiol.*, 49:42-53.
65. N. Arisue, Sanchez, L.B., Weiss, L.M., Muller, M., and T. Hashimoto (2002) Mitochondrial-type hsp70 genes of the amitochondriate protists, *Giardia intestinalis*, *Entamoeba histolytica* and two microsporidians. *Parasitology Internatl.*, 51:9-16.
66. H. Nozaki, M. Matsuzaki, M. Takahara, O. Misumi, H. Kuroiwa, M. Hasegawa, T. Shin-i, Y. Kohara, N. Ogasawara, and T. Kuroiwa (2003) The phylogenetic position of red algae revealed by multiple nuclear genes from mitochondria-containing eukaryotes and an alternative hypothesis on the origin of plastids. *J. Mol. Evol.*, 56:485-497
67. Nozaki, H., Matsuzaki, M., Misumi, O., Kuroiwa, H., Hasegawa, M., Higashiyama, T., Shin-i, T., Kohara, Y., Ogasawara, N. and Kuroiwa, T. (2004) Cyanobacterial genes transmitted to the nucleus before divergence of red algae in the Chromista. *J. Mol. Evol.*, 59: 103-113.
68. N. Arisue, Y. Maki, H. Yoshida, A. Wada, L.B. Sanchez, M. Muller, and T. Hashimoto (2004) Comparative analysis of the ribosomal components of the hydrogenosome-containing protest, *Trichomonas vaginalis*. *J. Mol. Evol.*, 59:59-71.
69. *N. Arisue, M. Hasegawa, and T. Hashimoto (2005) Root of the eukaryote tree as inferred from combined maximum likelihood analyses of multiple molecular sequence data. *Mol. Biol. Evol.*, 22: 409-420.
70. *Y. Nishimoto, N. Arisue, S. Kawai, A.A. Escalante, T. Horii, K. Tanabe, and T. Hashimoto (2007) submitted.

Virus evolution

71. *T.-K. Seo, J.L. Thorne, M. Hasegawa, and H. Kishino (2002) A viral sampling design for testing the molecular clock and for estimating evolutionary rates and divergence times. *Bioinformatics*, 18:115-123.
72. *T.-K. Seo, J. L. Thorne, M. Hasegawa, and H. Kishino (2002) Estimation of effective population size of HIV-1 within a host: a pseudomaximum-likelihood approach. *Genetics*, 160:1283-1293.
73. C. Sugimoto, M. Hasegawa, A. Kato, H.-Y. Zheng, H. Ebihara, F. Taguchi, T. Kitamura, and Y. Yogo (2002) Evolution of human polyomavirus JC: Implications for the population history of humans. *J. Mol. Evol.*, 54:285—297.
74. C. Sugimoto, M. Hasegawa, H.-Y. Zheng, V. Demenev, Y. Sekino, K. Kojima, T. Honjo, H. Kida, T. Hovi, T. Vesikari, J.A. Schalken, K. Tomita, Y. Mitsunobu, H. Ikegaya, N. Kobayashi, T. Kitamura, and Y. Yogo (2002) JC virus strains indigenous to Northeastern Siberians and Canadian Inuits are unique but evolutionally related to those distributed throughout Europe and Mediterranean areas. *J. Mol. Evol.*, 55: 322—335.
75. H.-Y. Zheng, C. Sugimoto, M. Hasegawa, N. Kobayashi, A. Kanayama, A. Rodas, M. Mejia, J. Nakamichi, J. Guo, L.G. Berthiaume, T. Kitamura, and Y. Yogo (2003) Phylogenetic relationships among JC virus strains in Japanese/Koreans and native Americans speaking Amerind or Na-Dene. *J. Mol. Evol.*, 56: 18-27.
76. Y. Zhang, N. Zheng, P. Hao, Y. Cao, and Y. Zhong (2005) A molecular docking model of SARS-CoV S1 protein in complex with its receptor, human ACE2. *Comput. Biol. Chem.*, 29:254-257
77. *Y. Nishimoto, T. Takasaka, M. Hasegawa, H.Y. Zheng, Q. Chen, C. Sugimoto, T. Kitamura, and Y. Yogo (2006) Evolution of BK virus based on complete genome data. *J. Mol. Evol.*, 63:341-352.
78. H.-Y. Zheng, Y. Nishimoto, Q. Chen, M. Hasegawa, S. Zhong, H. Ikegaya, N. Ohno, C. Sugimoto, T. Takasaka, T. Kitamura, and Y. Yogo (2007) Relationships between BK virus lineages and human populations. *Microbes and Infection* (in press)

Plant evolution

79. Y. Zhong, Q. Zhao, S. Shi, Y. Huang, and M. Hasegawa (2002) Detecting evolutionary rate heterogeneity among mangroves and their close terrestrial relatives. *Ecol. Lett.*, 5: 427-432.
80. W. Martin, T. Rujan, E. Richly, A. Hansen, S. Comelsen, T. Lins, D. Leister, B. Stoebe, M. Hasegawa, and D. Penny (2002) Evolutionary analysis of *Arabidopsis*, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proc. Natl. Acad. Sci. USA*, 99: 12246-12251
81. *Y. Nishimoto, O. Ohnishi, and M. Hasegawa (2003) Topological incongruence between nuclear and chloroplast DNA trees suggesting hybridization in the urophyllum group of the genus *Fagopyrum*

(Polygonaceae). *Genes Genet. Syst.*, 78: 139-153.

Others

82. O. Rickards, C. Martinex-Labarga, J.K. Lum, G.F. De Stefano, R.L. Cann (1999) Mitochondrial DNA history of the Cayapa Amerinds of Ecuador: detection of additional founding lineages for the native American populations. *Amer. J. Hum. Genet.*, 65:519-530
83. W.S. Watkins, M. Bamshad, M.E. Dixon, B. Bhaskara Rao, J.M. Naidu, P.G. Reddy, B.V.R. Prasad, P.K. Das, P.C. Reddy, P.B. Gai, A. Bhanu, Y.S. Kusuma, J.K. Lum, P. Fischer, and L.B. Jorde (1999) Multiple origins of the mtDNA 9-bp deletion in populations of South India. *Amer. J. Phys. Anthro.*, 109:147-158.
84. H. Sugimura, T. Kohno, K. Wakai, K. Nagura, K. Genka, H. Igarashi, B.J. Morris, S. Baba, Y. Ohno, C.M. Gao, Z.Y. Li, J.D. Wang, T. Takezaki, K. Tajima, T. Varga, T. Sawaguchi, J.K. Lum, J.J. Martinson, S. Tsugane, T. Iwamasa, K. Shinmura, and J. Yukota (1999) hOGG1 Ser326Cys polymorphism and lung cancer susceptibility. *Cancer Epidemiology Biomarkers and Prevention*, 8:669-674.
85. A. Kaneko, J.K. Lum, J. Yaviong, Y. Takahashi, T. Ishizaki, L. Bertilsson, T. Kobayashi, and A. Bjorkman (1999) High and variable frequencies of CYP2C19 mutations: medical consequences of poor drug metabolism in Vanuatu and other Pacific islands. *Pharmacogenetics*, 9:581-590.
86. J.K. Lum and R.L. Cann (2000) mtDNA lineage analyses: origins and migrations of Micronesians and Polynesians. *Amer. J. Phys. Anthro.*, 113:151-168.
87. Weisler, M., J.K. Lum, S. Collins, and W. Kimoto (2000) Status, health, and ancestry of a late prehistoric burial from Kwajalein Atoll, Marshall Islands. *Micronesica*, 32:191-220.
88. H.Y. Handoko, J.K. Lum, G. Rismalia, H. Kartapradja, A.S.M. Sofro, and S Marzuki (2000) Length variations in the COII-tRNA Lys intergenic region of mitochondrial DNA in Indonesian populations. *Human Biology*, 73:205-223.
89. *T. Akasaki, M. Nikaido, K. Tsuchiya, S. Segawa, M. Hasegawa, and N. Okada (2006) Extensive mitochondrial gene arrangements in coleoid Cephalopoda and their phylogenetic implications. *Mol. Phylogenet. Evol.* 38: 648-658.

Appendix2 Schedule of External Review

Place: Meeting Room of the Institute of statistical Mathematics

Date: March 27, 2007

12:30 – 12:40 Introduction by Genshiro Kitagawa (Director General)

- * Objective of evaluation

- * Terms of reference

12:40 – 13:30 lunch time

13:30 – 13:40 Preliminary meeting by Yoshiyasu Tamura (Vice Director General)

- * Selection of chairperson

- * Schedule and method

13:40 – 16:00 Meeting (members of the committee)

- * Discussion on evaluation

- * Arrangement for making review report

THE INSTITUTE OF STATISTICAL MATHEMATICS

TOKYO
JAPAN

ISM 分子進化研究プロジェクト

外部レビュー委員会レポート

2008 年 1 月

最 終 レ ポ ー ト

Contents

序文	19
0. 推奨事項の要約	20
1. 委員会とその作業	21
1.1 外部評価委員会	21
1.2 分子系統学研究グループ	21
1.3 評価項目	22
1.4 評価手順と方法	22
2. 学術的貢献：紹介ならびに賛辞	23
2.1 統計手法：成果とその影響力	24
2.2 分子進化：成果とその影響力	25
2.3 共同研究活動	26
3 社会への貢献	27
4 次世代の育成	27
5 将来の発展への展望および助言	27
5.1 推奨事項	28
付録 1: ISM 分子系統学研究グループの発表論文一覧	30
付録 2: 外部評価スケジュール	35

序文

統計数理研究所は平成 16 年 4 月に大学共同利用機関法人情報・システム研究機構の一研究所として再編された。これに先立ち、平成 15 年度には今後の統計数理の発展の方向を考慮して、「予測と発見」のためのモデリングや推論アルゴリズムなどの研究および統計ソフトウェアの開発を、ゲノム科学や地球・宇宙科学の具体的課題の解決を目指す組織として予測発見戦略研究センターを設置した。設立時は

- ・ ゲノム解析研究グループ
- ・ 動的磁気圏研究グループ
- ・ 地震予測研究グループ

の 3 グループによってスタートしたが、平成 17 年度には新たに 1 グループを加えるとともに

- ・ ゲノム解析グループ
- ・ データ同化グループ
- ・ 地震予測解析グループ
- ・ 遺伝子多様性解析グループ

に再編した。

このようなプロジェクト研究に関しては、自ら立てた目標、計画をどの程度まで実現したかの評価が不可欠である。今後順次、4 グループの中間評価、プロジェクト終了時の評価を実施していく予定であるが、その第 2 弾として平成 18 年度に外国人研究者 3 人、日本人研究者 1 人からなる外部評価委員会を構成し、ゲノム解析グループの中間評価を実施した。評価委員会に先立ち評価委員には、評価のための資料を送付した。

評価委員会の委員長であるデーヴィッド・ペニー教授（マッセイ大学、ニュージーランド）を中心にまとめられた英文の報告書は平成 19 年 7 月上旬に届けられた。添付している報告書の和訳は、民間翻訳会社により納入された和訳原文を統計数理研究所により専門的表現を監修したものである。諸事情により、評価報告書の公開が遅れたことお詫びしたい。

評価は評価者および評価を受けるものにとって多大の負担を伴うが、評価結果を将来の改善に活用してはじめてその苦勞が報われることになる。統計数理研究所は研究グループのメンバーとともに、今回の評価結果を真摯に受け止め今後の改善に繋げたいと考えている。

最後になるが、この評価のために貴重な時間と労力を割いてくださった評価委員の方々には厚くお礼を申し上げたい。

平成 20 年 1 月 31 日

統計数理研究所長
北川 源四郎

セクション 0 推奨事項の要約

当委員会は分子進化 (Molecular Evolution) グループが長期間にわたり挙げられた成果に感銘を受けており、ゲノムデータの更なる拡大に伴ってこの重要な分野における高度な統計技法への要求がますます高まるものと考えている。統計数理研究所 (Institute of Statistical Mathematics) において現在活動中のグループはデータ解析法としての尤度法を分子系統樹と系統進化の分野に現実的手法として導入したリーダーであり、急速な発展を遂げつつあるこの分野において日本はもとより世界的に重要な役割を果たしてきた。このような観点から当委員会は以下の事項を推奨する：

1. 分子進化に関する分野でグループを先導する研究者を新たに任命し、その募集を国際的に行う。
2. グループの研究領域を拡張し、統計計算法への比重を従来よりも大きくする。
3. グラフィックインターフェイスの経験を有するプログラマーをグループに含めることにより、ソフトウェアをより生物学者にとって扱い易いものにする。
4. グループの研究範囲を拡大して統計的生物情報学も包含するものとする。
5. 博士号取得後の研究者を増員する。

セクション 1 委員会と作業

1.1: 外部評価委員会 (名字のアルファベット順)

David Penny,	Massey 大学教授、ニュージーランド (議長)
Andrew J Roger,	Dalhousie 大学教授、カナダ
Hiddenori Tachida	九州大学教授、日本
Ziheng Yang,	University College London 教授 (FRS)、英国

1.2 分子系統学研究グループ

2007 年 3 月時点での構成員

Masami Hasegawa	教授	- 2007 年 3 月
Tetsuo Hashimoto	助教授	- 2002 年 3 月
	客員教授	2005 年 4 月 - 2007 年 3 月
Jun Adachi	学位取得研究員	1999 年 6 月 - 2000 年 4 月
	助教授	2003 年 2 月 -
Ying Cao	学位取得研究員	1998 年 4 月 - 2002 年 12 月
	助手	2003 年 1 月 -
Hiddenori Nishihara	学位取得研究員	2006 年 4 月 - 2007 年 3 月
Yuriko Nishimoto	大学院生	2001 年 12 月 - 2003 年 9 月
	学位取得研究員	2003 年 10 月 - 2007 年 3 月
Takahiro Yonezawa	大学院生	2002 年 4 月 - 2007 年 3 月
Atsushi Matsui	大学院生	2004 年 9 月 - 2007 年 3 月

1998 年のレビュー後に在籍した旧構成員

Hidetoshi Shimodaira	助手	1996 年 7 月 - 2002 年 3 月
Peter J. Waddell	学位取得研究員	1997 年 3 月 - 1999 年 3 月
Andrew Shedlock	学位取得研究員	1998 年 3 月 - 2000 年 11 月
Jeffrey Koji Lum	学位取得研究員	1998 年 3 月 - 2000 年 3 月
Margueritte Butler	学位取得研究員	1998 年 3 月 - 2000 年 2 月
Christopher Austin	学位取得研究員	1998 年 3 月 - 2000 年 3 月
David A. McClellan	学位取得研究員	1999 年 3 月 - 2001 年 2 月
Tal Pupko	学位取得研究員	2000 年 11 月 - 2002 年 9 月
Masato Nikaido	学位取得研究員	2003 年 4 月 - 2006 年 3 月
Rissa Ota	大学院生	1998 年 4 月 - 2000 年 9 月
Tae-Kun Seo	大学院生	1999 年 4 月 - 2002 年 9 月
Nobuko Arisue	大学院生	1999 年 4 月 - 2002 年 9 月
Yumie Murata	大学院生	2003 年 4 月 - 2005 年 3 月

1.3 評価項目

委員会には以下の項目に関する報告が求められた：

1. 以下の項目に関連した学術的寄与
 - (a) 統計手法：成果とその影響力
 - (b) 分子進化：成果とその影響力
 - (c) 共同研究活動
2. 社会への貢献
3. 次世代の育成
4. 将来の発展への展望および助言

1.4 評価手順と手法

委員会は 2007 年 3 月の 26 日および 27 日に研究所を訪問した。この訪問時に我々は以下の活動を行い、以下の情報が我々に提供された。

過去および現在の活動を説明する詳細な報告書が Masami Hasegawa 教授によって準備されていた。公表論文の詳細リストを本レポートの末尾に添付する。

3 月 26 日：同日午後に研究施設内において「分子進化の新たな地平線 – 統計科学としての分子系統学」をテーマとするシンポジウムが開催された。シンポジウム開催中は同グループのメンバーがそれぞれの成果を発表した。

3 月 27 日：外部レビュー委員会のメンバーが午前中にプレゼンテーションを行った。

それに続いて研究所長 Genshiro Kitagawa 博士および副所長 Yoshiyasu Tamura 博士との面会が行われた。午後には分子進化グループメンバーとさらに話をかわす機会が与えられ、また東京大学の Kishino 教授も研究所を訪問された。

このようにしてレビューパネルは必要とする情報すべてに触れることができた。

委員会のメンバーはこのようなレビューに参加する機会を与えてくれた Kitagawa 教授に感謝の意を表しますとともに、我々と懇談して頂いた同研究所のすべての研究員の方々、特に分子進化グループのメンバーの皆様に御礼申し上げ、訪問中に受けた数々の厚意に謝意を表します。また、レビュー資料が注意深くかつ完全に仕上げられており、レビューの実施前に配布されたことについても高く評価しております。

セクション 2 学術的貢献：紹介ならびに賛辞

我々が注目したのは 1998 から 2007 年に至る期間（前回の評価以降の期間）であるが、折りしも Masami Hasegawa 教授の退官とも機を一にして過去 25 年におよぶ分子進化グループの活動に敬意を表する好機であった。この期間全体にわたり同グループは世界的なリーダーであり続け、おそらくはすべての研究分野をとおして日本の研究グループとして最も世界的に注目を浴びて高く評価されたグループであった。たとえば、Hasegawa, Kishino and Yano (1985) および Kishino and Hasegawa (1989) の 2 つの論文はそれぞれが各種研究文献に 2000 回以上も引用されている。革新的かつ優れたこの統計手法の研究は過去 8 年間継続されて何報かの Shimodaira-Hasegawa 論文として公表されている。

過去 20 年間にわたり分子進化グループは分子進化と系統学という新しく開発されて急速に発展をとげつつある分野で大きな貢献をしてきた。同グループの貢献はデータ解析へ適用する統計的手法の開発およびその手法を配列データに適用して何種類かの主要な生物グループの発生学的関係解明に取り組むという両面に及んでいる。

上で触れたようなヌクレオチド置換のマルコフ過程モデルは同グループの科学者によって 1984 年と 1985 年に提案されて現在では HKY モデルとして広く知られており、DNA 配列進化モデルとして最も広く使用されている。同グループによって 1989 年に提唱された Kishino-Hasegawa 検定は複数の系統樹を最尤法と比較する手法として最も広く用いられている。Kishino、Miyata、および Hasegawa によるタンパク系統樹推定の最尤法はアミノ酸配列解析、特に生命の初期進化の研究で頻繁に用いられている。Adachi 博士を中心として開発されたプログラムパッケージ MOLPHY はすでに多くのユーザーを獲得しているが、それが再び更新作業中であることを知ったことも喜ばしいことであった。モデル選択の手法として赤池の情報量規準 (Akaike Information Criterion) を分子進化と系統学の研究者集団に導入したのも Hasegawa 博士のグループであった。

評価を担当する我々自身も分子進化と体系学の現役の研究者であることから、Hasegawa 博士のグループの研究がこの分野の正に先端を行くものであることに疑いを持たない。同グループは分子系統学に最尤法を適用したパイオニアである。その適用にあたって、同グループは分子進化学研究が統計的アプローチを取り入れるのを促進する重要な貢献を行った。同グループは日本のみならず世界を見渡しても最尤法および一般化最小二乗法を正面から取り上げた最初の重要な分子進化研究グループであり、系統樹のような複雑なデータ構造にこれらの手法を適用する能力を持ったグループである。前にも述べたように、同グループは DNA 変化に重要な確率過程モデル (Hasegawa-Kishino-Yano モデル、すなわち HKY モデル) を導入し、系統学に付随する統計的不確定性の研究のためにブートストラップおよび関連した検定法の使用を探索した。このような努力は思弁的議論や ad hoc データ解析に満ちていた当該分野を強固な理論的基盤を持つ応用統計が適用される分野へ変貌させるために重要な役割を果たした。

この分野で発生する統計的問題には複数のモデル（系統樹トポロジ）が関与し、それらは高度な対称性を持つ、しかし複雑な方法で相互に関連している。ツリーモデル空間は離散パラメータと連続パラメータが混在しているという点で特異な空間であるため既存の統計手法を適用できるために重要な正則性条件を満たしていない。それと同時に、これらすべての確率モデルはよく調べられている指数分布族に含まれる。偉大な統計学者 Jerzy Neyman (1971) がつとに思い描いていたごとく、進化の分子論的研究は新たな統計問題を生み出す肥沃な土壌である。

Neyman J. 1971. Molecular studies of evolution: a source of novel statistical problems. Pp. 1-27 in Gupta SS, and Yackel J, eds. Statistical decision theory and related topics. Academic Press, New York.

数多くの理論面での開発以外にも同グループは人間を始めとする生物システムに関する幾つかの重要な貢献を行ってきた。その例として挙げられるのは人類の系統発生、現代人類の進化と類人猿の関係、哺乳類と脊椎動物の間に存在する関係、さらには真核生物の起源などの大昔に遡る系統学的分岐の研究などが含まれ、最後の項目の研究には生命の樹の根底を探る非常に革新的な（複製遺伝子を使用する）アプローチが含まれていた。影響は単に方法論的な問題を越えて、同グループが分子論的な方法で多数の性質の異なる進化への疑問に対するエネルギー的な取り組みが生物学に多大な影響を及ぼしてきた。同グループが世界で最良かつ（理論と応用の）最もバランスのとれた研究グループであることについては検証できることである。

研究領域は進化論的な関係に統計手法を適用することから始まったが、現在では生物学と医学の広範な分野に応用が広がってきており、同グループが与えた影響は生物学のみならず分子進化の方法論開発にも及んでいる。同グループの全体としての研究は、人類の起源と進化から哺乳類の発生と進化を含む生物学の幾つかの領域に影響を及ぼし、さらには命のツリーの原初分岐、生命体自体の発生に続く最初の分岐にまで及んでいる。同グループはこれらの領域内でのみ有効であったばかりではなく、強固な基盤を持つ統計手法を導入したことにより同グループの影響ははるかに広い範囲に及んでいる。

献辞の最後に指摘したいポイントは同グループが高度の影響力を持つジャーナルに論文を公表し、科学文献に広く引用されていることである。たとえば、同グループは次に示す 3 種類の学際的ジャーナルに論文を発表している。

学際科学；

ジャーナル名	インパクトファクター	分野におけるランク
Science	30.93	1
Nature	29.27	2
Proc. Natl Acad. Sci. USA	10.23	3

Source: Journal Citation Reports 2005
Institute for Scientific Information, Philadelphia, PA

このようなジャーナルへの発表は多くの研究者が追い求めているものである。

これに加えて、同グループは進化生物学の基礎研究で最も大きな影響力を持つ 2 種類のジャーナル、すなわち Systematic Biology（インパクトファクター 10.33）および Molecular Biology and Evolution（インパクトファクター 6.23）にも論文を発表しており、これらの研究は日本のみならず世界中で広く引用されている。簡単にまとめると、系統学という新しい分野と進化の数多くの重要問題に基礎的統計技法を適用し、高度のインパクトを持つ論文をジャーナルに発表することにより分子進化グループは 25 年以上にわたり世界のリーダーであり続けてきた。同グループの成果に献辞を捧げられるのは我々の喜びとするところである。以上で献辞を終了し、前回の外部レビュー以降の 8 年間を更に詳細に検討してみる。

セクション 2.1 統計手法：成果とその影響力

分子進化グループは系統樹の再構築への尤度法の適用、特に系統発生論上の推定を検定するための統計手法の開発で重要な貢献をしてきた。系統発生はそれぞれが独自のパラメータを持つモデルとして捉えるのが良い方法であり、したがって異なる系統発生は異なるパラメータ空間に存在する。そのため、最尤推定の漸近的特性を利用する従来型の理論を適用することができず、推定モデルすなわち系統発生の不確定性を評価するのは困難な課題となっている。D.R. Cox は枝分かれを伴わない異なるモデルを比較する何種類かの検定法について議論したが、これらの方法が広く用いられることはなかった。

分子進化グループの Kishino と Hasegawa は 1989 年に 2 つの系統樹の対数尤度計算値間の差の正規近似に基づく K-H 検定を導入した。ただしこの検定法は 3 つ以上の系統樹を比較する多重検定には適用できなかった。1999 年に至り、Shimodaira と Hasegawa は多重検定に対応する補正を加えた S-H 検定を導入した。Shimodaira はまた殆ど偏りのない検定法 (AU 検定) も導入した。これらの論文は現在では分子系統発生の分野で広く使用され、数千もの論文によって引用されている。S-H 検定と AU 検定は今回のレビューの対象期間中に分子進化グループによって発表されたものであることを特記しておきたい。これらの検定法を持つ一般性と、モデル選択に関連して統計学の論文が現在寄せている注目を考えるならば、これらの研究は統計学においても重要な貢献を成すものである。

同グループは系統樹の検定以外にも分子進化と系統発生の統計学的な問題に大きな寄与をしている。同グループは分子時計と緩和分子時計モデルを種分岐の時代同定に応用した最初のグループの 1 つであり、この問題については現在も方法論的な前進を遂げつつある (Seo et al., 2002; Hasegawa et al. 2003)。さらに最近ではゲノム規模での大型データセットがますます一般的になりつつある。同グループではゲノムデータセットを解析するための各種戦略の評価を行っており (Nishihara et al. 2007)、その結果は大型データセットを取り扱う他の科学者達へも貴重な洞察力を提供している。

セクション 2.2 分子進化：成果とその影響力

ここでは同グループが開発した方法の生物学の諸問題への応用に注目して解説する。まず最初に述べたいのは、同グループ内で理論と応用が相互にサポートし、刺激し合ってきたことに我々が感銘を受けたことである。この相互作用により現実の生物学的問題に対する強力な統計手法の応用が生まれてきた。たとえば、Hasegawa と Cao は東京工業大学の Okada グループと共同で配列データを SINES からの配列データと組み合わせる新しい方法を開発しており、この手法はますます強力なものになりつつある。SINES (short interspersed nuclear elements) はゲノム全体にランダムに挿入可能な要素であるが、異なる種の隣接する DNA 領域の配列を決定することによって他と重複しない挿入であることを確認することができる。挿入サイトに隣接する領域の配列を決定することによって共有挿入を識別できるため、その挿入サイトがすべての種で同じであるか否かを確認することができる。哺乳類のツリートポロジは現在では SINES によって決定が可能であり、したがって配列を独立に使用して分岐の量 (したがって時代) を推定することができる。単純節約法はこのようなデータの最尤推定量である (Appl. Math. Lett. 17: 785 2004)。我々は最尤法の枠内で SINES と配列データを公式にしたがって組み合わせる改良法の出現を期待しており、その時期は近いと考えている。

Hasegawa と Hashimoto、および共同研究者達は真核生物の初期進化に重要な貢献をしてきたが、その目的は細胞を含む最も単純な核相互の関係を明らかにすることであった。これらの原稿には紅藻類の系統発生と、紅藻核ゲノムの組成に対する葉緑体内部共生の遺伝的関与を説明した多数の重要な論文が含まれている (紅藻、特に「のり」 (*Porphyra yezoensis*) は日本において特別な経済的重要性を持っている)。純粋科学的

観点から見てさらに注目すべきなのは Arisue、Hasegawa および Hashimoto が最近

(2005) 公表した論文であり、この中で著者らは多数の遺伝子に「別々な」最尤法解析を適用して真核生物系統樹の根源決定を試みた。この原稿は解析の系統誤差同定に用いた手法が極めて高度であることから注目される。この問題は議論の絶えない困難かつ重要な問題であるが、上記論文はこれに取り組んだ最近の解析の 1 つである。

近年、分子進化グループはウィルス集団の遺伝子と系統発生の分野でも重要な貢献を果たしている。特に RNA ウィルスは非常に迅速な進化を遂げるため進化の過程と、その結果としての疾病の症状を理解するための格好のモデルである。ウィルスの分野で同グループは分子時計検定のためのサンプリング設計、祖先集団サイズの擬似最尤推定量の両方で新しい方法論を提供したばかりでなく、ウィルス進化過程を利用して人類進化の追跡とウィルスの基本的分子生物学の理解に寄与する重要な生物学的発見をした。さらに同グループの活動は生物学の他の領域もカバーしており、その中でも最も重要な 1 つと考えられるのは古代の人類移動パターンを人類集団の遺伝子の多様性を利用して解明しようとする試みである。

これまでに説明した主題指向的な研究トピックスに加えて、グループの各メンバーはその折々で多様な単発的共同プロジェクトにも参画しており、それらのプロジェクトが採り上げるトピックスは分子進化の極めて広い範囲をカバーしている。採り上げられたトピックスの数例のみを列挙すると人類の多様性の研究、ガンの生物学、寄生虫の進化とミトコンドリア遺伝子などが含まれる。分子進化グループの知的環境が刺激に満ちていることは明らかであり、これが多様でしかも相互に補完的な生産的研究プログラムを育てている。これらについては続く「共同研究活動」のセクションで説明する。最後に、同グループが以前作成したコンピュータープログラムの幾つかを更新中であることは喜ばしいことである。アミノ酸レベルでの研究が可能になることにより幾つもの利点もたらされる。

セクション 2.3 共同研究活動

前のセクションでも説明したように、ISM グループは他のグループとの連携にも積極的であった。これまでのセクションで触れた原稿の多くは ISM の統計分子進化研究者と日本および世界の他の生物学者との生産的共同作業の賜物である。これら外部の研究者達は明らかに ISM 分子進化グループが持つ独自の専門知識を求めたのであり、グループが非常に高い評価を受けていることを裏付けている。

外部研究者との広範な共同作業を示す最も直接的な方法はグループが論文を共同発表した研究所や組織を列挙することである。日本国内では何ヶ所かの研究所（基礎生物学研究所 [岡崎]を含む）や博物館のほかに総合研究大学院大学のいくつかの他の部門が含まれ、大学としては東京大学、京都大学、東京工業大学、早稲田大学（東京）、大阪市立大学、名古屋大学、広島大学が含まれる。このような日本国内における共同研究の広範なリスト（その幾つかは何年にもわたって継続している）は DNA データの統計解析ばかりではなく生物学的意味付けの理解についても同グループの力量が求められていることを端的に示すものである。ISM の分子進化部門はこのように日本国内において確固たるリーダーシップを発揮している。

海外へ目を向けると、共同研究の期間という点では日本の幾つかの事例がより長く継続しているとは言え、共同研究リストはさらに広がりを見せる。アジア地域から始めると中国科学院（武漢）と Kyungpook National University（韓国）との共同研究がある。欧州ではドイツ、英国およびイスラエルとの間で共同研究が行われた。ニュージーランドでは Otago および Massey 大学、および北米では数例を挙げただけでも Boston 大学、

California 大学 (Berkeley)、Michigan 大学などが含まれる。以上から得られる結論は、分子進化グループが外部の研究者から共同研究を求められるような卓越かつ重要な研究を行ってきたということであり、同時に ISM の同グループが常に外へ目を向けていたということである。分子進化グループが常に外部の動きに注目していたことがアイデアを他の研究者に伝達するのに役立ったのである。

セクション 3 社会への貢献

科学者が社会へ成し得る最も重要な貢献は革新的研究を実施すること、および次世代の研究者を育成することである。これまでに詳しく説明したとおり、この意味においても同グループは社会への貢献に大きな成功を収めている。貢献が持つ特別な側面として、分野の異なる科学者や一般大衆に科学的な研究結果を知らせることも重要である。同グループは主として革新的な研究の実施と次世代を担う若い科学者の育成に重点を置いてきたが、後者の側面に対しても貢献してきている。

本稿の対象期間においても、Hasegawa 教授は生物学の異なる分野の研究者を対象とする 2 冊の日本語の著作 (共著) を上梓している他に客員教授として他大学で多数の講義を行っている。これに加えて、以前同グループに属していたメンバーが日本の主要大学で教鞭をとっている。さらに、Hasegawa 教授はこの期間中に一般大衆を対象とする講演を何回か行っている。革新的研究を主眼とする研究所の性質を考えるならば、貢献の社会的側面という意味でも同グループは優れた成果を挙げたと言うことができる。

セクション 4 次世代の育成

このセクションでは特に博士号取得後の特別研究員に独自の研究経験を積ませる指導と大学院教育について注目する。我々の委員会に課せられた課題は主として同グループの研究活動の評価であるが、同グループが大学院教育に果たしている役割にいくらか言及するのも意味のあることであると考ええる。我々の見るところ、我々が予期したほどには多数の大学院生が同グループの存在を生かしていないように思われる。ISM 分子進化グループは非常に強力であることから、日本の大学院生達はこのグループから学び、彼らの勉学の基礎として利用すべきである。6 月初めに中国を訪問した外部レビュー委員会の 3 名のメンバーが当地で見聞した PhD 教育増強への強力な後押しも我々の考えを更に強固なものにした。総合研究大学院大学の設立は重要な前進ではあるが、ISM がより多くの大学院生を受け入れることが強く望まれる。

セクション 5 将来の発展への展望および助言

ここまで我々は分子進化グループの過去の研究について説明してきたが、それに加えてこの卓越したグループが将来においてより大きな影響力を発揮できる研究方向について触れておくのも意味のあることであると考ええる。最初に言及したいのは統計やコンピューター科学から生ずる新たな機会を常に考慮する必要がある点であり、特にゲノムデータは巨大化しており、今後も拡張し続けられると思われる。この中でも特に衝撃的であるのは最後に挙げたゲノムデータである。新技術の導入に伴ってゲノム DNA 配列の奔流がさらに増大し続けるのはほぼ確実である。洪水のようなこのデータを現実の生物学的理解に結びつけるためには統計および計算方法の更なる改良が必須であり、これらの改良には生物学者との積極的な連携が伴わなければならない。

(1) 計算分子生物学との更に密な連携。我々が見たとおり、現在のグループは分子進化の実験的研究およびこの分野への新しい統計技法の応用という点で極めて強力である。しかし、成長著しい分野である計算分子生物学（生物情報学の一部）との接続という点では十分な努力が払われていないと思われる。系統発生を統計的に解析するコンピュータープログラムが Adachi と Hasegawa (MolPhy パッケージ) によって開発されて配布されているが、この数年の間に計算分子生物学への注目がコンピューター科学者の間で急速に高まりつつあることを考慮すれば、今が東京地域のコンピューター科学者との連絡を密にする好機であると考えられる。混合配列 (combined 'alignment') や系統樹推定問題などの多様な興味深い問題が存在することから、同グループは日本のコンピューター科学者がこれらの興味深い問題を取り上げる刺激として機能できると思われる。多他分野のコンピューター科学者が抱える困難の 1 つはこれらの問題を統計的に捉えようとする意志の不足である。同グループが統計研究所に属している事実が影響を及ぼして日本のコンピューター科学者はこれらの問題をアルゴリズムの問題としてばかりではなく統計の問題として捉えるのに寄与すると思われる。

(2) 分子進化と統計理論の研究なより密な連携。
Shimodaira 博士（あるいは過去における Kishino 教授）のような優れた統計学者の存在が同グループの研究の大きな助けとなってきた。将来においても強力な統計的要素が必須であり、研究主題全体についてもこの要素が非常に重要であると考えられる。その際、将来の統計手法はデータ解析のニーズに対応したものでなければならないことに注意が必要である。統計手法に関する研究が次第にデータ解析のニーズから乖離する傾向を示す可能性も存在するが、このような傾向の発生は避けなければならない。このような調整を行う 1 つの可能な方法として、Shimodaira 博士が指導する統計学専攻の大学院生に分子進化データ解析の実際的重要性に関わる問題に取り組ませることが考えられる。

セクション 5.1 推奨事項

当委員会は分子進化グループの過去 25 年にわたる仕事から大きな感銘を受けた。長年にわたりグループリーダーを勤められた Hasegawa 教授が退官されるにあたり、ISM がこの発展中の分野で継続して世界のリーダーとして地位を保つための助言をさせていただく好機とも考える。このため、当委員会は以下の推奨事項を取りまとめた。これらの推奨事項が同グループ発展のために寄与することを願っている。

1. 分子進化の基礎的な統計学を担当する新たな上級研究員を任命する。この募集は国際的に行われるものとする。

当委員会はグループを更新するための優れた核はすでに存在すると考え、より強力な統計的手法を実現し、かつユーザーコミュニティのニーズに対応した新しい機会（計算と膨大化する DNA 配列データの両面から）を取り入れ易くするために幾らかの配置替えを許すのが妥当と考える。そのためには強固な統計理論を背景に持ち、グループ全体を新たな挑戦に導くことのできる新たなリーダーが必要と考える。この地位に魅力を感じる日本人候補者ももちろん存在するであろうが、原則としてこの募集は国籍にとらわれず国際的に行われるべきであると考えられる。海外に居住する日本人研究者が応募することも可能であるが、任命が日本国内からなされた場合であっても新しいグループディレクターに国際的な募集の結果選任されたという自信を与えることになるであろう。

2. グループの研究領域を拡張し、統計計算法への比重を従来よりも大きくする。

新しい解析方法はより広い範囲のツリーとモデルを探索するため相当量の計算機時間を必要とする。この結論が導かれる背景についてはこのセクション（セクション 5）の導入部でも説明しておいた。取り扱うデータセットがますます巨大化しているため、モデルに追加パラメータを検討するのが妥当な時期に達している – これにより生物学的実態を取り込むのが容易になる。研究者達は今まで無かった疑問を問いかけており、その中には遺伝子のどのサイトが正の淘汰の対象となるのかという問題も含まれている。これに加えて、データに含まれる追加的な調節シグナルを推測できる蓋然性も大きくなってきている。これにより、例えば遺伝子内の情報が複数レイヤの制御を経て如何にして細胞を機能させる情報へ変換されるのかをより優れた方法で推定できるようになる。

3. グラフィカルユーザーインターフェイス（GUI）の経験を持つプログラマーをグループに含める。

当グループから見ると、グループの才能ある研究者がソフトウェア（GUI）の実装やプラットフォーム依存性のない（Windows、Unix、Linux、MacOSX のいずれでも使用できる）アプリケーション開発という複雑な作業のために実に多大な時間を消費している。このような作業はそれ自体が複雑なものであり、このような作業を専門プログラマーに任せて研究者達が方法論的な革新や評価、実データセットでの検証に振り向けられる時間を解放するのが良いと考える。この項目は 2 番とも明らかに関係しており、少数のプログラマーだけで全部の仕事をこなすことはできない。

4. グループの研究範囲を拡大して統計的生物情報工学も包含する。

遺伝子配列がより完全に決定されるようになるに伴い、より優れた統計解析の必要性が増大している。この概念はセクション 5 の導入部でも述べたことであり、この推奨項目は項目 2 および 3 の内容を引き継ぐものである。生物学者は自分のラボにおける実験を自由に設計できるという意味で非常に革新的であり得るが、新しいタイプの解析を自由に行えるほど革新的な統計学者ではない場合が多い。今後 20 ないし 30 年にわたり、生物学者と共同で遺伝子からより多くの情報を引き出す能力を備えた統計学者への需要が非常に大きいことは確実である。すなわち、生物学者のみでは研究が進められないことが確実である。ここに拡大の一途を辿るのが確実である機会が存在する。例を 1 つだけ挙げるとすれば、ある特定の突然変異が病徴を引き起こすか否かは殆ど分かっていない – 膨大に存在する未使用の遺伝子情報を利用して革新的な統計的思考を展開できる殆ど無限の空間がここに存在するのである。

5. 博士号取得後の研究者の数を増やす。

分子進化という主題は、統計推測の複雑な一面を示すという意味で大学院生の教育にまたとないテーマである。当委員会から見て ISM は大学院教育にとって理想的な場所を提供するものであり、また大学院教育を後押しするという観点からも新たにグループディレクター（就職活動をサポートする）という地位を設けるのが望ましい。理想的には日本人と日本人以外の学生両方の増員を希望するが、後者については政府の政策に依存する部分が大きい。世界的に見ると、国外から PhD 学生を呼び寄せられる能力をもってその研究機関を評価し、政府が研究機関の質を担保しようとするトレンドが見られる。ここでは分子進化が大学院レベルの学生教育に最適なトピックス（学生が広い範囲を定量的に考察し続けなければならない）であること、および統計数理研究所（ISM）がこのプロセスに理想的な位置を占めることを再び述べるに止めておく。

付録 1 :

ISM 分子系統学研究グループの発表論文一覧 1999 – 2007

* の付いた論文は論文集に収録されている。

統計手法

1. *H. Shimodaira and M. Hasegawa (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol. Biol. Evol.*, 16:1114-1116.
2. *J. Adachi, P. Waddell, W. Martin and M. Hasegawa (2000) Plastid genome phylogeny and a model of amino acid substitution for proteins encoded by chloroplast DNA. *J. Mol. Evol.*, 50:348—358.
3. *R. Ota, P.J. Waddell, M. Hasegawa, H. Shimodaira and H. Kishino (2000) Appropriate likelihood ratio tests and marginal distributions for evolutionary tree models with constraints on parameters. *Mol. Biol. Evol.*, 17:798—803.
4. D. Penny and M. Hasegawa (2001) Covarion model of molecular evolution. In S. Brenner and J.H. Miller, eds., *Encyclopedia of Genetics*, pp. 473—477, Academic Press (San Diego).
5. H. Kishino and M. Hasegawa (2001) Maximum likelihood. In S. Brenner and J.H. Miller, eds., *Encyclopedia of Genetics*, pp. 1157—1160, Academic Press (San Diego).
6. *H. Shimodaira and M. Hasegawa (2001) CONSEL: a program for assessing the confidence of phylogenetic tree selection. *Bioinformatics* 17: 1246-1247.
7. H. Shimodaira (2001) Multiple comparisons of log-likelihoods and combining nonnested models with applications to phylogenetic tree selection. *Comm. Statist. Theory Methods*, 30:1751-1772.
8. D.A. McClellan and K.G. McCracken (2001) Estimating the influence of selection on the variable amino acid sites of the cytochrome *b* protein functional domains. *Mol. Biol. Evol.*, 18:917-925.
9. *H. Shimodaira (2002) An approximately unbiased test of phylogenetic tree selection. *Syst. Biol.*, 51:492-508.
10. *T. Pupko, D. Huchon, Y. Cao, N. Okada, and M. Hasegawa (2002) Combining multiple data sets in a likelihood analysis: Which models are the best? *Mol. Biol. Evol.*, 19:2294-2307.
11. *T. Pupko, I. Pe'er, M. Hasegawa, D. Graur, and N. Friedman (2002) A branch-and-bound algorithm for the inference of ancestral amino-acid sequences when the replacement rate varies among sites: Application to the evolution of five gene families. *Bioinformatics*, 18: 1116-1123
12. *T. Pupko, R. Sharan, M. Hasegawa, R. Shamir, and D. Graur (2003) Detecting excess radical replacements in phylogenetic trees. *Gene* 319: 127-135.
13. P.J. Waddell, H. Mine, and M. Hasegawa (2004) INTEROGATE 1.0. Exploration and testing of stationarity, reversibility and clock-likeness in sequence data. *Computer Science Monograph (ISM)*, 31:1-22.
14. *H. Shimodaira and M. Hasegawa (2005) Assessing the uncertainty in phylogenetic inference. In R. Nielsen, ed., *Statistical Methods in Molecular Evolution*, pp.463—493, Springer.
15. *K. Shimizu, J. Adachi, and Y. Muraoka (2006) ANGLE: A sequencing errors resistant program for predicting protein coding regions in unfinished cDNA. *J. Bioinf. Comput. Biol.*, 4:649-664.
16. Z. Li, Q. Liu, M. Song, Y. Zheng, P. Nan, Y. Cao, G. Chen, Y. Li, and Y. Zhong (2005) Detecting correlation between sequence and expression divergences in a comparative analysis of human serpin genes. *BioSystems* 82:223- 226.

脊椎動物の進化

17. *P. Waddell, N. Okada, and M. Hasegawa (1999) Towards resolving the interordinal relationships of placental mammals. *System. Biol.*, 48:1-5.
18. *P. Waddell, Y. Cao, J. Hauf, and M. Hasegawa (1999) Using novel phylogenetic methods to evaluate mammalian mtDNA, including amino acid-invariant sites-LogDet plus site stripping, to detect internal conflicts in the data, with special reference to the positions of hedgehog, armadillo, and elephant. *System. Biol.*, 48: 31-53.
19. D. Penny, M. Hasegawa, P.J. Waddell, and M.D. Hendy (1999) Mammalian evolution: Timing and implications from using the LogDeterminant transform for proteins of differing amino acid composition. *System. Biol.*, 48:76-93.
20. *P. Waddell, Y. Cao, M. Hasegawa, and D.P. Mindell (1999) Assessing the Cretaceous superordinal

- divergence times within birds and placental mammals by using whole mitochondrial protein sequences and an extended statistical framework. *System. Biol.*, 48:119-137.
21. *D.P. Mindell, M.D. Sorenson, D.E. Dimcheff, M. Hasegawa, J. Ast, and T. Yuri (1999) Interordinal relationships of birds and other reptiles based on whole mitochondrial genomes. *System. Biol.*, 48:138-152.
 22. *Y. Cao, K.S. Kim, J.H. Ha and M. Hasegawa (1999) Model dependence of the phylogenetic inference: Relationship among carnivores, perissodactyls and cetartiodactyls as inferred from mitochondrial genome sequences. *Genes Genet. Syst.*, 74: 211-217.
 23. C.C. Austin (1999) Lizards took express train to Polynesia (1999) *Nature*, 397: 113-114
 24. C.C. Austin (1999) Island colonization by *Lipinia noctua* (Reptilia: Scincidae) in Melanesia: Molecular phylogeny and population structure based on mitochondrial cytochrome b and 12S rRNA Genes. *Island Herpetofauna: Origin, Current Status and Conservation* (ed. T. Ota), Elsevier Press, pp. 169—189.
 25. C.C. Austin (1999) Colonization, isolation and speciation of lizards on Pacific islands: a molecular genetic perspective. *Aquabiology*, 21:313-318.
 26. C.C. Austin and G. Zug (1999) Molecular and morphological evolution in the south-central Pacific skink *Emoia tongana* (Reptilia: Squamata): uniformity and human-mediated dispersal. *Austral. J. Zool.*, 47: 425-437.
 27. Y. Tohyama, H. Kasama-Yoshida, M. Sakuma, Y. Kobayashi, Y. Cao, M. Hasegawa, H. Kojima, Y. Tamai, M. Tanokura, and T. Kurihara (1999) Gene structure and amino acid sequence of Latimeria chalumnae (coelacanth) myelin DM20: phylogenetic relation of the fish. *J. Neurochem.*, 24:867—873.
 28. *J.K. Lum, M. Nikaido, M. Shimamura, H. Shimodaira, A.M. Shedlock, N. Okada, and M. Hasegawa (2000) Consistency of SINE insertion topology and flanking sequence tree: Quantifying relationships among cetartiodactyls. *Mol. Biol. Evol.*, 17:1417-1424.
 29. *M. Nikaido, M. Harada, Y. Cao, M. Hasegawa and N. Okada (2000) Monophyletic origin of the order Chiroptera and its phylogenetic position among Mammalia, as inferred from the complete mitochondrial DNA sequence of a Japanese megabat, the Ryukyu flying fox (*Pteropus dasymallus*). *J. Mol. Evol.*, 51:318-328.
 30. *Y. Cao, M.D. Sorenson, Y. Kumazawa, D.P. Mindell and M. Hasegawa (2000) Phylogenetic position of turtles among amniotes: evidence from mitochondrial and nuclear genes. *Gene*, 259:139-148.
 31. *Y. Cao, M. Fujiwara, M. Nikaido, N. Okada and M. Hasegawa (2000) Interordinal relationships and time-scale of eutherian evolution as inferred from mitochondrial genome data. *Gene*, 259:149-158.
 32. A.M. Shedlock, M.C. Milinkovitch, and N. Okada (2000) SINE evolution, missing data, and the origin of whales. *Syst. Biol.* 49:808-817.
 33. A.M. Shedlock and N. Okada (2000) SINE insertions: powerful tools for molecular systematics. *BioEssays*, 22:148-160.
 34. M.A. Butler, T.W. Schoener, and J.B. Losos (2000) The relationship between habitat type and sexual size dimorphism in Greater Antillean Anolis lizards. *Evolution*, 54:259-272.
 35. Y. Tohyama, T. Ichimiya, H. Kasama-Yoshida, Y. Cao, M. Hasegawa, H. Kojima, Y. Tamai, and T. Kurihara (2000) Phylogenetic relation of lungfish indicated by the amino acid sequence of myelin DM20. *Mol. Brain Res.*, 80:256—259
 36. M. Onuma, Y. Cao, M. Hasegawa and S. Kusakabe (2000) A close relationship of Chiroptera with Eulipotyphla (core Insectivora) suggested by four mitochondrial genes. *Zool. Sci.*, 17:1327—1332.
 37. S. Ohdachi, N.E. Dokuchaev, M. Hasegawa, and R. Matsuda (2001) Intraspecific phylogeny and geographic variation of six species of northern Asiatic *Sorex* shrews based on the mitochondrial cytochrome b sequences. *Mol. Ecol.* 10:2199—2213
 38. M. Sumida, Y. Kanamori, H. Kaneda, Y. Kato, M. Nishioka, M. Hasegawa, and H. Yonekawa (2001) Complete nucleotide sequence and gene rearrangement of the mitochondrial genome of the Japanese pond frog *Rana nigromaculata*. *Gene Genet. Syst.* 76:311—325
 39. *M. Nikaido, F. Matsuno, H. Hamilton, R.L. Brownell Jr., Y. Cao, D. Wang, Z. Zhu, A.M. Shedlock, R.E. Fordyce, M. Hasegawa, and N. Okada (2001) Retroposon analysis of major cetacean lineages: The monophyly of toothed whales and the paraphyly of river dolphins. *Proc. Natl. Acad. Sci. USA*, 98: 7384-7389.
 40. *M. Nikaido, K. Kawai, Y. Cao, M. Harada, S. Tomita, N. Okada, and M. Hasegawa (2001) Maximum likelihood analysis of the complete mitochondrial genomes of eutherians and a reevaluation of the phylogeny of bats and insectivores. *J. Mol. Evol.* 53:508-516
 41. K. Kawai, M. Nikaido, M. Harada, S. Matsumura, L.-K. Lin, Y. Wu, M. Hasegawa, and N. Okada (2002) Intra- and interfamily relationships of Vespertilionidae inferred by various molecular markers including SINE insertion data. *J. Mol. Evol.*, 55: 284-301

42. M.A. Butler and J.B. Losos (2002) Multivariate sexual dimorphism, sexual selection, and adaptation in Greater Antillean Anolis lizards. *Ecological Monographs*. 72:541-559.
43. *M. Nikaido, Y. Cao, N. Okada, and M. Hasegawa (2003) The phylogenetic relationships of insectivores with special reference to the lesser hedgehog tenrec as inferred from the complete sequence of their mitochondrial genome. *Genes Genet. Syst.* 78: 107-112.
44. *M. Hasegawa, J.L. Thorne, and H. Kishino (2003) Time scale of eutherian evolution estimated without assuming a constant rate of molecular evolution. *Genes Genet. Syst.* 78: 267-283
45. *Y. Murata, M. Nikaido, T. Sasaki, Y. Cao, Y. Fukumoto, M. Hasegawa, and N. Okada (2003) Afrotherian phylogeny as inferred from complete mitochondrial genomes. *Mol. Phylogenet. Evol.* 28: 253-260.
46. *M. Nikaido, Y. Cao, M. Harada, N. Okada, and M. Hasegawa (2003) Mitochondrial phylogeny of hedgehogs and monophyly of Eulipotyphla. *Mol. Phylogenet. Evol.* 28: 276-284.
47. K. Kawai, M. Nikaido, M. Harada, S. Matsumura, L.-K. Lin, Y. Wu, M. Hasegawa, and N. Okada (2003) The status of the Japanese and East Asian bats of the genus *Myotis* (Vespertilionidae) based on mitochondrial sequences. *Mol. Phylogenet. Evol.*, 28: 297-307.
48. A.M. Shedlock, T.W. Pietsch, M.G. Haygood, P. Bentzen, M. Hasegawa (2004) Molecular systematics and life history evolution of anglerfishes (Teleostei: Lophiiformes): Evidence from mitochondrial DNA. *Steenstrupia*, 28:129-144.
49. *T. Sasaki, M. Nikaido, H. Hamilton, M. Goto, H. Kato, N. Kanda, L.A. Pastene, Y. Cao, R.E. Fordyce, M. Hasegawa, and N. Okada (2005) Mitochondrial phylogenetics and evolution of Mysticete whales. *Syst. Biol.*, 54: 77-90.
50. Butler M.A. (2005) Foraging mode of the chameleon, *Bradypodion pumilum*: a challenge to the sit-and-wait versus active forager paradigm? *Biol. J. Linnean Soc.*, 84:797-808
51. *A. Kurabayashi, C. Usuki, N. Mikami, T. Fujii, H. Yonekawa, M. Sumida, and M. Hasegawa (2006) Complete nucleotide sequence of the mitochondrial genome of a Malagasy poison frog *Mantella madagascariensis*: Evolutionary implications on mitochondrial genomes of higher anuran groups. *Mol. Phylogenet. Evol.* 39: 223-236
52. *H. Nishihara, M. Hasegawa, and N. Okada (2006) Pegasoferae, an unexpected mammalian clade revealed by tracking ancient retroposon insertions. *Proc. Natl. Acad. Sci. USA* 103: 9929-9934
53. H. Endo, T. Yonezawa, F. Rakotondrapary, M. Sasaki, and M. Hasegawa (2006) The adaptational strategies of the hindlimb muscles in the Tenrecidae species including the aquatic web-footed tenrec (*Limnogale mergulus*). *Annals of Anatomy* 188:383-390
54. S.D. Ohdachi, M. Hasegawa, M.A. Iwasa, P. Vogel, T. Oshida, L.-K. Lin and H. Abe (2006) Molecular phylogenetics of soricid shrews (Mammalia) based on mitochondrial cytochrome *b* gene sequences: with special reference to the Soricinae. *J. Zool.*, 270:177-191 (Errata 270:199-200).
55. *T. Sasaki, M. Nikaido, S. Wada, T.K. Yamada, Y. Cao, M. Hasegawa, and N. Okada (2006) *Balaenoptera omurai* is a newly discovered baleen whale that represents an ancient evolutionary lineage. *Mol. Phylogenet. Evol.* 41:40-52
56. *M. Watanabe, M. Nikaido, T.T. Tsuda, T. Kobayashi, D. Mindell, Y. Cao, N. Okada, and M. Hasegawa (2006) New candidate species most closely related to penguins. *Gene* 378: 65-73.
57. *M. Munemasa, M. Nikaido, S. Donnellan, C.C. Austin, N. Okada, and M. Hasegawa (2006) Phylogenetic analysis of diprotodontian marsupials based on complete mitochondrial genomes. *Genes Genet. Syst.* 81:181-191.
58. L.A. Pastene, M. Goto, A.N. Zerbini, K. Watanabe, Y. Bessho, D. Kerem, N. Kanda, M. Hasegawa, R. Nielsen, F. Larsen, and P.J. Palsbøll (2007) Marine speciation during periods of global warming: timing of origin and radiation of the common minke whale, *Balaenoptera acutorostrata*. *Molecular Ecology* (in press)
59. *A. Matsui, F. Rakotondrapary, M. Hasegawa, and S. Horai (2007) Determination of a complete lemur mitochondrial genome from feces. *Mammal Study* (in press)
60. *T. Yonezawa, M. Nikaido, N. Kohno, Y. Fukumoto, N. Okada, and M. Hasegawa (2007) Molecular phylogenetic study on the origin and evolution of Mustelidae. *Gene* (in press)
61. *H. Nishihara, N. Okada, and M. Hasegawa (2007) Rooting the eutherian tree—The power and pitfalls of phylogenomics (submitted).

真核生物の進化

62. M. Hasegawa and T. Hashimoto (1999) Phylogenetic position of amitochondriate protists in the evolution of

eukaryotes. *Biol. Bull.*, 196: 389—392.

63. *N. Arisue, T. Hashimoto, J.A. Lee, D.V. Moore, P. Gordon, C.W. Sense, T. Gaasterland, M. Hasegawa, and M. Muller (2002) The phylogentic position of the pelobiont *Mastigamoeba balamuthi* based on sequences of rDNA and translation elongation factors EF-1 α and EF-2. *J. Euk. Microbiol.*, 49:1-10.
64. *N. Arisue, T. Hashimoto, H. Yoshikawa, Y. Nakamura, G. Nakamura, F. Nakamura, T. Yano, and M. Hasegawa (2002) Phylogenetic position of *Blastocystis hominis* and of stramenopiles inferred from multiple molecular sequence data. *J. Eukaryot. Microbiol.*, 49:42-53.
65. N. Arisue, Sanchez, L.B., Weiss, L.M., Muller, M., and T. Hashimoto (2002) Mitochondrial-type hsp70 genes of the amitochondriate protists, *Giardia intestinalis*, *Entamoeba histolytica* and two microsporidians. *Parasitology Internatl.*, 51:9-16.
66. H. Nozaki, M. Matsuzaki, M. Takahara, O. Misumi, H. Kuroiwa, M. Hasegawa, T. Shin-i, Y. Kohara, N. Ogasawara, and T. Kuroiwa (2003) The phylogenetic position of red algae revealed by multiple nuclear genes from mitochondria-containing eukaryotes and an alternative hypothesis on the origin of plastids. *J. Mol. Evol.*, 56:485-497
67. Nozaki, H., Matsuzaki, M., Misumi, O., Kuroiwa, H., Hasegawa, M., Higashiyama, T., Shin-i, T., Kohara, Y., Ogasawara, N. and Kuroiwa, T. (2004) Cyanobacterial genes transmitted to the nucleus before divergence of red algae in the Chromista. *J. Mol. Evol.*, 59: 103-113.
68. N. Arisue, Y. Maki, H. Yoshida, A. Wada, L.B. Sanchez, M. Muller, and T. Hashimoto (2004) Comparative analysis of the ribosomal components of the hydrogenosome-containing protest, *Trichomonas vaginalis*. *J. Mol. Evol.*, 59:59-71.
69. *N. Arisue, M. Hasegawa, and T. Hashimoto (2005) Root of the eukaryote tree as inferred from combined maximum likelihood analyses of multiple molecular sequence data. *Mol. Biol. Evol.*, 22: 409-420.
70. *Y. Nishimoto, N. Arisue, S. Kawai, A.A. Escalante, T. Horii, K. Tanabe, and T. Hashimoto (2007) submitted.

ウィルスの進化

71. *T.-K. Seo, J.L. Thorne, M. Hasegawa, and H. Kishino (2002) A viral sampling design for testing the molecular clock and for estimating evolutionary rates and divergence times. *Bioinformatics*, 18:115-123.
72. *T.-K. Seo, J. L. Thorne, M. Hasegawa, and H. Kishino (2002) Estimation of effective population size of HIV-1 within a host: a pseudomaximum-likelihood approach. *Genetics*, 160:1283-1293.
73. C. Sugimoto, M. Hasegawa, A. Kato, H.-Y. Zheng, H. Ebihara, F. Taguchi, T. Kitamura, and Y. Yogo (2002) Evolution of human polyomavirus JC: Implications for the population history of humans. *J. Mol. Evol.*, 54:285—297.
74. C. Sugimoto, M. Hasegawa, H.-Y. Zheng, V. Demenev, Y. Sekino, K. Kojima, T. Honjo, H. Kida, T. Hovi, T. Vesikari, J.A. Schalken, K. Tomita, Y. Mitsunobu, H. Ikegaya, N. Kobayashi, T. Kitamura, and Y. Yogo (2002) JC virus strains indigenous to Northeastern Siberians and Canadian Inuits are unique but evolutionally related to those distributed throughout Europe and Mediterranean areas. *J. Mol. Evol.*, 55: 322—335.
75. H.-Y. Zheng, C. Sugimoto, M. Hasegawa, N. Kobayashi, A. Kanayama, A. Rodas, M. Mejia, J. Nakamichi, J. Guo, L.G. Berthiaume, T. Kitamura, and Y. Yogo (2003) Phylogenetic relationships among JC virus strains in Japanese/Koreans and native Americans speaking Amerind or Na-Dene. *J. Mol. Evol.*, 56: 18-27.
76. Y. Zhang, N. Zheng, P. Hao, Y. Cao, and Y. Zhong (2005) A molecular docking model of SARS-CoV S1 protein in complex with its receptor, human ACE2. *Comput. Biol. Chem.*, 29:254-257
77. *Y. Nishimoto, T. Takasaka, M. Hasegawa, H.Y. Zheng, Q. Chen, C. Sugimoto, T. Kitamura, and Y. Yogo (2006) Evolution of BK virus based on complete genome data. *J. Mol. Evol.*, 63:341-352.
78. H.-Y. Zheng, Y. Nishimoto, Q. Chen, M. Hasegawa, S. Zhong, H. Ikegaya, N. Ohno, C. Sugimoto, T. Takasaka, T. Kitamura, and Y. Yogo (2007) Relationships between BK virus lineages and human populations. *Microbes and Infection* (in press)

植物の進化

79. Y. Zhong, Q. Zhao, S. Shi, Y. Huang, and M. Hasegawa (2002) Detecting evolutionary rate heterogeneity among mangroves and their close terrestrial relatives. *Ecol. Lett.*, 5: 427-432.
80. W. Martin, T. Rujan, E. Richly, A. Hansen, S. Comelsen, T. Lins, D. Leister, B. Stoebe, M. Hasegawa, and D. Penny (2002) Evolutionary analysis of *Arabidopsis*, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proc. Natl. Acad. Sci. USA*, 99: 12246-12251
81. *Y. Nishimoto, O. Ohnishi, and M. Hasegawa (2003) Topological incongruence between nuclear and

chloroplast DNA trees suggesting hybridization in the urophyllum group of the genus *Fagopyrum* (Polygonaceae). *Genes Genet. Syst.*, 78: 139-153.

その他

82. O. Rickards, C. Martinex-Labarga, J.K. Lum, G.F. De Stefano, R.L. Cann (1999) Mitochondrial DNA history of the Cayapa Amerinds of Ecuador: detection of additional founding lineages for the native American populations. *Amer. J. Hum. Genet.*, 65:519-530
83. W.S. Watkins, M. Bamshad, M.E. Dixon, B. Bhaskara Rao, J.M. Naidu, P.G. Reddy, B.V.R. Prasad, P.K. Das, P.C. Reddy, P.B. Gai, A. Bhanu, Y.S. Kusuma, J.K. Lum, P. Fischer, and L.B. Jorde (1999) Multiple origins of the mtDNA 9-bp deletion in populations of South India. *Amer. J. Phys. Anthro.*, 109:147-158.
84. H. Sugimura, T. Kohno, K. Wakai, K. Nagura, K. Genka, H. Igarashi, B.J. Morris, S. Baba, Y. Ohno, C.M. Gao, Z.Y. Li, J.D. Wang, T. Takezaki, K. Tajima, T. Varga, T. Sawaguchi, J.K. Lum, J.J. Martinson, S. Tsugane, T. Iwamasa, K. Shinmura, and J. Yukota (1999) hOGG1 Ser326Cys polymorphism and lung cancer susceptibility. *Cancer Epidemiology Biomarkers and Prevention*, 8:669-674.
85. A. Kaneko, J.K. Lum, J. Yaviong, Y. Takahashi, T. Ishizaki, L. Bertilsson, T. Kobayashi, and A. Bjorkman (1999) High and variable frequencies of CYP2C19 mutations: medical consequences of poor drug metabolism in Vanuatu and other Pacific islands. *Pharmacogenetics*, 9:581-590.
86. J.K. Lum and R.L. Cann (2000) mtDNA lineage analyses: origins and migrations of Micronesians and Polynesians. *Amer. J. Phys. Anthro.*, 113:151-168.
87. Weisler, M., J.K. Lum, S. Collins, and W. Kimoto (2000) Status, health, and ancestry of a late prehistoric burial from Kwajalein Atoll, Marshall Islands. *Micronesica*, 32:191-220.
88. H.Y. Handoko, J.K. Lum, G. Rismalia, H. Kartapradja, A.S.M. Sofro, and S Marzuki (2000) Length variations in the COII-tRNA Lys intergenic region of mitochondrial DNA in Indonesian populations. *Human Biology*, 73:205-223.
89. *T. Akasaki, M. Nikaido, K. Tsuchiya, S. Segawa, M. Hasegawa, and N. Okada (2006) Extensive mitochondrial gene arrangements in coleoid Cephalopoda and their phylogenetic implications. *Mol. Phylogenet. Evol.* 38: 648-658.

付録 2： 外部評価スケジュール

場所：統計数理研究所会議室

日時：2007 年 3 月 27 日

12:30 – 12:40 所長の挨拶と概要説明

- ・評価の目的
- ・評価項目

12:40 – 13:30 昼食

13:30 – 13:40 準備会議

- ・委員長選出
- ・スケジュールと評価方法の検討

13:40 – 16:00 委員の会合

- ・評価の検討
- ・評価報告書についての相談

大学共同利用機関法人
情報・システム研究機構
統計数理研究所

〒106-8569 東京都港区南麻布 4 - 6 - 7

Tel 03-5421-8706 (総務課庶務係)

Fax 03-3443-3552

URL <http://www.ism.ac.jp/>

分子進化
研究プロジェクト
外部評価委員会
報告書

2008年2月

大学共同利用機関法人 情報・システム研究機構
統計数理研究所