Dear Mr Code...

Author(s): Jouni Issakainen
Reviewed work(s):
Published by: International Association for Plant Taxonomy (IAPT)
Stable URL: http://www.jstor.org/stable/1224439
Accessed: 01/06/2012 14:39

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at http://www.jstor.org/page/info/about/policies/terms.jsp

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

*International Association for Plant Taxonomy (IAPT)* is collaborating with JSTOR to digitize, preserve and extend access to *Taxon.*
Dear Mr Code ...  

J. Issakainen¹

In this letter to the imaginary “Mr Code”, some of the handicaps of the current International code of botanical nomenclature are discussed with respect to consideration of molecular data. Problems and solutions are outlined to central questions such as type material, anamorphic taxa, publishing, description and diagnosis, ranks of taxa, role of genera and species, and accepted number of correct names. The provisional concept “basonomy” is proposed for simultaneous classification of multilocal molecular characters.

Dear Mr Code,

Thank you very much for your kind letter of 1993 from Tokyo (Greuter & al., 1993). You and your co-workers have done an excellent job in setting practical guidelines to classifying living organisms. You have mastered a task which, because of the immense number of taxa, deserves true respect.

Since time passes and you are surely preparing for your next message on improving taxonomy, I trust you do not mind if I share with you some informal thoughts that your letter inspired. These have emerged during a few years when I, having the background of a field biologist and medical mycologist, have had the opportunity to visit the area of molecular systematics. Being a kind of borderline walker I feel myself rather incompetent as compared to the real experts in these fields, including nomenclature. Therefore, I hope you understand that the ideas which I present below do not constitute a ready model but are preferentially meant as a source of inspiration.

1. Type material. – My first impression of your letter is that you greatly value type material as a physical reference point for the names of taxa. I fully agree with you. However, I think that this valuable principle has not been sufficiently applied to all parts of your writing.

Principle II and Articles 7-10 and 37 fix the names to physical, dead type specimens or, in some cases, to figures. From a molecular point of view, however, these kinds of materials differ totally from each other, and neither of them is satisfactory. In spite of the rapid development in the field, deoxyribonucleic acid (DNA) which is traceable to the described species is very difficult to extract from a figure. Even if we talk about real herbarium specimens, methods which rely on DNA are, in practice, incomplete. They may require masses of homogeneous, uncontaminated material, for instance. Few dead type specimens are pure and profuse enough for future

¹ Mycology and Parasitology Laboratory, Turku University Central Hospital, P. O. Box 52, FIN-20521 Turku, Finland.
studies on the entire genome. This applies especially to the older specimens which are prioritised by you in Principle III.

In the future, I hope you could suggest how to get rid of printed type material, and this as soon as possible. If and when such a transition is made, would it not be trouble-saving to change directly to the better type specimens described below, so to allow full-scale molecular studies?

A future type specimen should resemble the traditional one as much as possible, and perhaps the older type should constitute a part of the specimen. However I foresee that a proper type must, in the future, be a purposefully planned combination of profuse extant dead material, ex-type cultures and molecular products such as isolated DNA which can also be conserved. Formal support to such a development could be given by expanding the use of living and molecular epitypes (Article 9.7) to become a routine for all typification, not only for problematic cases. Guidelines and institutional frameworks for optimal preservation of each organism group should be systematically created.

2. Anamorph taxa. – As a second problem, it appears that you do not allow the description of holomorph taxa without teleomorphic elements (Article 59), although the lack of meiospore production is often a secondary feature and “form taxa” based on anamorphs may not reflect natural relationships.

A great number of molecular characters may be obtainable to prove that a given fully mitosporic form taxon is part of a known meiospore-forming taxon of a particular rank – or that it forms a separate taxon in a given rank. For instance, we may be sure that a permanently anamorphic or non-sporulating fungus is part of a given apothecia-forming genus (or family) and that it is fully separate from all known species of the genus (or from the known genera of the family).

Before any DNA-based conclusions regarding nomenclature can be drawn, each holomorphic taxon should be circumscribed by molecular sequencing and by other means. Work to this end is being pursued in several research teams around the world. We have now finally got the means to really penetrate into the secrets of non-sporulating taxa and surely it is unbearable if these tools cannot ultimately be applied to a unified nomenclature. But how to make it formal?

Once we know enough of a given taxon, could we not just amend its formal description (Article 47, *mutatis characteribus*) to include a base-by-base molecular definition of the taxon? After this, we could simply treat these features as “elements representing the teleomorph”, in the sense of Article 59.1. Gathering sufficient genomic data from an anamorphic species would, consequently, be equated with finding flowers of a plant or sexual spores of a fungus and this would allow transferring it to, or describing it in, a suitable holomorph genus. International expert committees could act as the fora which would discuss and decide as to the suitable genomic areas, form of presentation and the extent of the data needed. Once there was general agreement on these “molecular flowers”, it would be logical that nomenclature also should permit giving final names to genera which totally lack sporulation. “Final” implies here that the characters themselves will prove to be durable.

3. Literature handling. – Mr Code, I can see that one of your hobbies must be literature, especially older literature. Your Principles III and IV and Article 11 equate the correct name of a taxon with priority of publication. Article 6.1 and Articles 29-
31 further define “effective publication” by specifying the allowed techniques of printing and transmitting the taxonomic results.

Having worked in a botanical library myself, I also like the smell of dusty leather covers and antique prints. However, each devotion has its costs. Your principles, once necessitated by the limits of information technology, imply that a vast amount of poorly organised taxonomic literature of varying quality should be completely known. Practical handicaps of data retrieval considered, the taxonomist is condemned to use an excessive amount of time for literature searches and technical paper handling, at the expense of primary observations. You do not accept electric delivery of information, although this would be essential, especially with molecular data.

I am glad to learn that your team is actively engaged in developing standardised taxonomy databases. I believe that even less shyness is in order. As a first step, the existing taxonomic (or, more widely, systematic) literature could be actively scanned for electronic retrieval by computers. It would still retain its paper-bound layout but could be searched based on terms such as taxon names or authors and downloaded to individual personal computers and printed. Depending on the effort put into the selection of literature at the retrieval stage, the database could become either a partial compilation or, preferably, an exhaustive set of all the publications that were internationally agreed to contain valid nomenclatural actions; other past publications could then be ignored.

As the second step, I suggest that all this refined literature should be retrievable from carefully planned and “intelligent” databases. In these databases the data should be broken down into its elements as far as possible, e.g., DNA base locations, standard deviations of spore length and speed of hyphal growth in given conditions. Observations from various disciplines, such as electron microscopy, biochemistry and in vitro drug studies could be directly plotted against morphological data and for further analysis. Interesting research questions would probably arise from improved connections between bits of data. Once sufficiently refined, morphological illustrations could ultimately be generated automatically from such a database.

In pace with this development, standards should be set for effective electronic publication of new taxa. Molecular and general biology data sets would gradually expand and allow a more flexible concept of describing and publishing taxa. One would not be bound to the structure of the modern scientific paper. Rather, researchers who by some standard would be qualified to submit data could send either single data units or extensive data sets to complement electronic databases. Such submissions would not be published or, indeed, publishable in modern journals. In future data networks, each used data unit could also be traced, as to reliability, and compensated for to institutions or individual scientists who produced it. These principles could be called “cellular responsibility” and “cellular copyright”, respectively.

Nomenclatural conclusions in the sense of your present views, Mr Code, could in the future be largely automated and be updated on a regular, even daily, basis if there is interest in the process of knowledge accumulation in the historical perspective. In cellular systems, the formal authorship of a given taxon need not be rigid but may be diluted to hundreds of workers in different research teams.

While I hail the benefits of the computer for everyday data handling I am not naive as to its vulnerability to damage. Therefore, rules should be established for distributing a sufficient amount of paper backups, as a continuation to modern journals.
4. Concepts of description and diagnosis. – In Article 32 you state that a valid name must be accompanied by a description or a diagnosis of the taxon. Since molecular characters will increasingly be used, these basic terms will need to be re-evaluated in the future.

A “diagnosis” includes only those characters of the taxon which the author finds that separate it from other taxa (Article 32.3). It is thus a definition of an abstract unit. The “description”, on the other hand, should contain the diagnosis (Recommendation 32B.1) but also a more extensive (preferably exhaustive?) account of the object’s features. Thus, if I understand this wording correctly, the nature of the “description” is a list of a large number of primary observations.

In morphology-based and identification-oriented taxonomy, these two categories of text may seem quite similar. However, their principal difference becomes apparent when molecular data are used and need to be combined with several faculties of science. In an extreme case, a molecular “diagnosis” of a taxon could consist of only one base-denoting letter, accompanied by a code number which defines the base’s genomic location. A full molecular “description”, on the other hand, would consist of billions of aligned letters, encompass decades of research work and be practically illegible and incomprehensible in paper format.

Moreover, between these two extremes, there is an infinite variety of possible lengths of a genomic account. This problem needs to be addressed in your future messages. The practical questions of a molecular taxonomy worker may have forms like: how long a sequence is allowed, in which gene, and in which organism group, for describing a taxon of which rank? The system should not allow changing the verbal taxon name every time a new base is added to the list. Cladistic classifications also depend on taxa selected as comparison points and this adds to the variation of names available.

What I fear in your present view is the very deep fixation of taxon borders into laconic verbal descriptions, especially when these deal with poorly defined morphological features. For the purposes of future classification, these may be much less relevant and less informative than specific base locations in a given sequenced genome area.

I believe that the Latin language (Article 36) will not be very useful in future descriptions. This follows from the assumption that the whole taxonomy will be more directly derived from large data sets rather than from verbal phrases. As a first step, Latin descriptions of taxa could be translated into a technically feasible language, such as taxonomic English enriched with Latin terms. This can also be largely automated during the above suggested data retrieval projects.

In the middle of this turmoil I trust that molecular systematists would have much to learn from your long experience of defining taxa. Although molecular classifications can be published based on calculations on sequence data, there is the trouble of defining concretely each taxon against all of its relatives. Demanding each author to point out a “taxonomic definition sequence” for each molecular taxon he or she publishes might increase the durability of names based on DNA sequences.

5. Artificiality of ranks. – I guess you have served some extra time in the army, Mr Code. You like ranks, don’t you? In Articles 3.1 and 4.1 you define universal, stepwise hierarchical ranks of taxa. In Article 35 you further state that a name of a taxon is not validly published unless its rank is clearly indicated. No name has pri-
ority outside the rank in which it was originally published (Article 11.2). Are you sure this is a durable standpoint, Mr Code?

Your ranking system implies that any taxonomic result which an author gets in a given taxon is, by definition, subordinated to his or her general vision of all existing organisms. But few taxonomists, I think, consider themselves as possessors of a clairvoyance about what a "final" ranking really is. If your system would be applied to the army, a night guard who often needs to report something should start every announcement by shouting the rank of the officer whom he guesses is hearing him into the darkness. Nobody answers, but if he guesses the rank wrongly, his report will just be silently ignored. This seems to be a lottery which is kept ongoing by the illusion of the existence of a durable hierarchy but which leads to an increasing pile of unnecessary synonyms.

The molecular data which keep on accumulating bring both good and bad news to your concept of ranking. As to the good news, they permit a more objective classification of all ranks, and they provide some hard evidence that can be used as a basis for establishing suprageneric ranks.

The bad news is that sequence-based dendrograms show that the evolution of species is a highly quantitative and relative process. Comparing ranks between adjacent clades is something artificial, man-made. Subdivisions allowed by Articles 4.2 and 4.3 postpone but do not solve this principal problem. The guesses of the night guard may become closer to the truth, but they are still guesses in a box game which has no equivalence in nature. During the decades to come, ranking is bound to change as data accumulate.

The emerging molecular data make us also increasingly aware of the unity of all organisms. If your suggestions would be taken seriously in future phylogenetic trees, new features found in a malaria plasmodium, for instance, might cause alterations in the universal rank definitions that would necessitate changes of rank, name and authority in every species of vascular plants. A taxonomic perpetuum mobile is thus foreseeable, and I believe that this solution cannot be the final one.

The principal problems of ranking could be postponed for some decades. Together with a general neo- or epitypification by recent specimens, a compromise classification could be created which utilises molecular definitions but applies the ranks and names of current taxonomy. Artificial baselines approximating modern ranks in different groups could be agreed on. The situation could then be frozen for some time by discouraging changes of ranks and subordinating these to an international committee. Further in the future I believe the taxonomic names in your sense must be more clearly separated from molecular and other detailed data.

6. Floating genera and species. – Your Article 2 defines the rank of species as basic to the taxonomic system. Article 23 further regulates that the binary name of a species must consist of the name of the genus, followed by a specific epithet. Thus, your so-called basic rank never appears alone but always as a member of another very central rank, the genus.

The binomial naming system leads directly to some artificial features of classification, such as unspecific genera or, especially in evolving families, oscillating generic limits. My main point is, however, that both of these primary ranks float in the space like clouds in the sky. They may have ambiguous borders and most often lack fixed points of reference as to which level of variation should be called "generic", which level "specific", or something between. While molecular studies are
providing us with sorely needed data for justified relative divisions, I doubt whether they offer any definite answer to the problems of defining your two basal ranks. More probably, the opposite will emerge.

When taxonomy is based on genetic sequences, the only conceivable basic rank is "two identical sequences", or full homology. Even this definition of identity, although seemingly clear, is valid only for a given genomic area. Specimens will be classified differently if different genomic loci are compared.

7. The number of correct names. – In Principle IV and Article 11.1 you state that each taxon can only bear one correct name. This implies that it can have only one correct place in the system of classification. However, the features of an organism are derived from the genome, and the evolutionary history of different loci of an organism's genome may vary.

Variation near the traditional species rank is caused by sexual recombination. This is complemented by different speeds of mutation in different populations and in different parts of the genome. But even in higher ranks, when prolonged mating barriers eventually break, the taxa recombine between ranks – a situation which a puritan taxonomist would consider highly inappropriate. In the long run, genes and gene fragments also duplicate and are reallocated inside a single cell line. Their biological tasks evolve, as well. I anticipate that molecular studies will reveal much more of genetic heterogeneity in all ranks and organism groups than we can currently imagine.

We easily forget that different parts of a single organism's genome may have a different evolutionary history. Your hope of a monolithic system seems to be programmed into man, as it were. Even in molecular systematics, emphasis is laid on the question as to which genes make "good" chronometers for the whole organism's evolutionary process. One is tempted to believe that once a good gene is found, the rest of them will be something secondary to taxonomy. However, the survival, mating and external appearance of species – all characters of good traditional holomorph taxonomy – are based on a sum effect of thousands of genes with their individual histories. It is also true that a mutation or recombination in a very small fraction of genome, totally independent from the clock genes, may be named in the traditional system as an infraspecific taxon such as a forma. A minute change could also lead to a technical mating obstacle and, eventually, to a new species.

Classical taxonomy as discussed by you is comparable with a totalitarian human society where a person can be a member of a single association only (family, club, party, nation, religion). According to this system, if a carrot plant is transfected with genes of a rabbit, it is still called a carrot; it has the membership card of the Carrot Club, exclusively. In contrast, a durable molecular taxonomy should, obviously, have a truly multi-local character. It should allow a single specimen to be simultaneously classified as belonging to several different phylogenetic groupings, depending on the contents of the studied genomic areas. This system could be compared with a modern liberal society where a person can acquire memberships of many associations, including ones with contradictory aims.

For purposes of discussion, I propose the term "basonomy" to describe a formal taxonomy which could be developed for classifying specimens according to their DNA base sequences. In such a system, every studied specimen has got a membership card in just as many classification systems ("basonomies") and classification units therein ("basons"), as it has gene areas ("basonomic loci") suitable to such
comparisons. Just as every traditional taxon should be given its “taxonomic definition sequence”, also each bason should be defined by a set of individual (perhaps not adjacent) DNA bases which comprise the full diagnosis of that bason. This could be called a “basonomic definition sequence”.

Future databases should allow classification of DNA features of any rank of taxa from the topmost domains to the very lowest, even momentary, genetical units. These include garden cultivars of plants, mould strains of different sexes, single sperm cells and melanocyte strains in different naevi of the skin of one’s back. If a given DNA or RNA molecule exists in a single cell as several slightly differing copies, basonomy will allow classifying each nucleic acid strand separately, at each moment of time, to the extent this is needed and methods are available. This kind of taxonomy would obviously be dependent on efficient computers but so are many other disciplines of science today. We conceive of the backside of the moon as a neutral fact, although computers are needed for planning a flight there.

The immense genomic variation is by no means only a burden. If genes have evolved at different speeds, a more reliable “stereoscopic view” of evolution might be obtained by comparing the phylogeny of several loci in several species. If ecological functions of these genes (such as resistance to drought, cold, or an aquatic environment) are known, different environmental eras in the taxon’s history could perhaps be timed in relation to the divergence of taxa.

8. Naming of molecular taxa. – In your system, the name of each species consists of two words, namely, the generic name and the specific epithet (Article 23.1). You also give some closer regulations and recommendations as to the explicit form of these words (Articles 23 and 60). These principles may need to be re-evaluated when the whole spectrum of genetic variation is named.

If the “basonomic” system is employed, the number of classification units would increase explosively. How could these be handled in a practical way? The clue might be to realise that names, DNA sequences, and Latin descriptions are all just queues of symbols. If clarity is achieved why do they need to be so different?

We could, for instance, agree that the full sequence of a basonomic locus (e.g. all 1000 first bases of the small subunit ribosomal DNA) is the “description” of the bason. The basonomic definition sequence (e.g. 37 bases scattered over different parts of the basonomic locus) could directly comprise the “diagnosis” of the bason. A small number of these, in turn, such as the three first base-denoting letters with their standardised location numbers, could be the “name” of the bason. Thus, an author would validly publish a bason and its name of any rank by publishing the relevant sequences in an agreed manner.

Characterising a given specimen by only three DNA bases implies that a coarse classification of the specimen is already known from the context. If this is not the case, a more complete bason name could be combined by adding step by step the three-letter “names” of every basonomic rank, from the highest to the smallest needed. Actually, this is what already is done when we add the family, order or class names, as needed for clarity. Used stepwise, the bason names also directly represent modern identification keys which are a significant part of a systematic article.

Epilogue. – Is all of this too radical? This is only the way to transfer the core of your long-tested naming principles, Mr Code, into the molecular world, while avoiding any unnecessary information. Basonomic code names would resist verbal
expression and remembrance, but computers bear it with ease. Actually, a separate mass-data handling system, something like basonomy, would relieve the pressures to change your own naming principles too soon. If you could only vigorously bind the modern nomenclature with the increasing molecular data, and simultaneously diminish unsubstantiated name changes, I cannot see why the verbal names could not continue as valuable indexing words long into the future.

Mr Code, in this critical letter, I have not yet mentioned some obvious facts: for centuries you and your colleagues have acted as mankind's eyes looking at the miraculous spectrum of life around – and inside – ourselves. In many cases, you have done this under most adverse conditions, often ignored by society, and you have been driven only by your own stamina and desire to understand nature.

If one does not name the organisms, one does not know of them. Such organisms would just be green spots in a jungle or, as is the case of millions of organisms even today, they would not exist for us at all. Due to the action of one species, Homo sapiens L., more and more of these forms of life are driven to extinction before they have been granted the first and smallest gesture of human recognition: given a name. Obviously, all who work in the field of biology are grateful to you for letting us get a glimpse of the spectrum of nature. I wish you every success and endurance in the decades to come.

In this letter I hope to have initiated some discussion about a more detailed naming which may be needed in the near future. However, I trust both of us understand that the scales we use, whatever their form, are not important. They are just our own, incomplete tools to approach the breathing, swarming and pulsating richness of life.

Yours sincerely,

Jouni Issakainen

Literature cited