Formation of prokaryote names from personal names: a review of current practice and a proposal to emend Appendix 9 of the International Code of Nomenclature of Prokaryotes

Mark J. Pallen[1,2,*]

Abstract

The practice of naming elements from the natural world after notable individuals stretches back to ancient times. This practice of creating eponyms—terms derived from personal names—has been carried forward into prokaryotic nomenclature, where the International Code of Nomenclature of Prokaryotes (ICNP) sets guidelines for creating scientific names from personal names. However, these guidelines can be seen as culturally biased, disjointed and, on occasion, misguided. Here, with the goal of modernizing these recommendations to render them more user-friendly, coherent and inclusive, I review current practice in the light of precedents and key linguistic and cultural principles, while questioning the applicability of the first-name/last-name paradigm for many cultural traditions. Procedural challenges include romanization of the personal name (including handling of diacritics), creation of a short and agreeable latinized stem, assignment of the stem to a declension and addition of suffixes or compound word components to create genus names or species epithets, customizing the approach for names and stems that end in a vowel. I review the pros and cons of stem augmentation, which involves addition of an extra 'i' to the original stem. Next, I formulate a coherent workflow, which I incorporate into a Python script to enable computer-based automation of name creation. Rather than following the ICNP in limiting discussion to a few dozen mainly European names, I examine how these principles work out when applied to the tens of thousands of last names under which scientists publish in the PubMed database, focusing on edge cases where conventional approaches fail, particularly very short and very long names. Drawing on these explorations and analyses, I propose emendations to the advice currently presented in the ICNP to usher in a modern, consistent, pragmatic and globally inclusive approach to the creation of prokaryotic eponyms.

DATA SUMMARY

Supplementary material Figshare link: https://figshare.com/s/aa95daebca9be767b6b2 [1].

INTRODUCTION

Since antiquity, components of the natural world have been named after notable individuals. Thus, in the works of Pliny the Elder, we find the plant *Artemisia* named after the goddess Artemis, *Paeonia* named after the mythical physician to the gods Paeon and *Euphorbia* named after the Greek physician Euphorbus [2]. This practice of creating *eponyms*—terms derived from personal names—has been carried forward into taxonomic nomenclature, where many species and genera of plants, animals and prokaryotes have been named after people [3–5].

The International Code of Nomenclature of Prokaryotes (ICNP) provides directions for formulating prokaryotic names from personal names in Appendix 9 [6]. There has been recent discussion over the value of eponyms in science, as they may be tainted by association with colonialism or other undesirable connotations [7]. While prokaryotic nomenclature has generally avoided such associations, the fact remains that most current prokaryotic eponyms are linked to males of European heritage. The guidance

within the ICNP reflects this bias in citing examples largely drawn from names of European origin (with a handful from Japan, but none at all from Africa, South Asia, China, Indonesia, Turkey or the Arab world), which sits uneasily with the global nature of modern microbiology. Furthermore, rather than offering a unified, consistent and easy-to-follow workflow, suitable for automation, the current guidance is presented as a disjointed set of unrelated processes, spread across four tables. In addition, some of the current guidance is misguided or even clearly inappropriate.

Here, with the goal of modernizing these recommendations, I first review current practice, exploring the linguistic and cultural principles at work when taxonomic names are created from personal names, collating and reverse engineering the existing guidance in the ICNP, while also examining precedents and recommendations from elsewhere in taxonomic nomenclature.

I then formulate a coherent workflow, with clear step-by-step instructions that can be followed by humans and computer programs. I subsequently incorporate these instructions into a Python script to enable computer-based automation of name creation [8, 9]. In so doing, I make the assumption—articulated in a previous publication [8]—that rather than create scientific names only as needed, it is preferable to generate the names *en masse* and make them available to the scientific community in advance of any particular use case.

Rather than following the ICNP in limiting discussions to a few dozen mainly European names, I examine how these principles apply to the tens of thousands of last names under which scientists publish in the PubMed database, focusing on the most commonly used names and edge cases where conventional approaches fail. I then revaluate the workflow and use the amended script to generate over a million eponyms suitable for use in prokaryotic nomenclature. Finally, drawing on these explorations and analyses, I close with proposed emendations to the advice currently presented in the ICNP, including coherent procedures and a more inclusive set of examples.

REVIEW OF CURRENT PRACTICE

Guiding principles

The ICNP states that it is good practice to ask the person to be honoured by a scientific name for permission if they are alive. In search of a more inclusive approach that respects personal and cultural differences, I suggest that the person to be honoured should be allowed to choose how their personal name is used. If that is not possible, the personal name should be used in a way that honours the cultural and linguistic conventions of the name bearer. How the personal name is used to form a scientific name also needs to take into account Principle 3 and Rule 6 of the Code—which make clear that scientific names of prokaryotic taxa must be Latin words or latinized words treated as Latin—and Recommendations 6(1) and 6(2), which advise authors to avoid creating names that are long or difficult to pronounce or otherwise disagreeable. In addition, any proposed scientific name must not have been used previously in prokaryotic, zoological or botanical nomenclature, which adds to the challenge of creating distinctive names.

There is no explicit guidance within the ICNP on the full set of the steps required to create a scientific name from a personal name. Much is left unsaid, assuming the reader already has a working knowledge of Latin grammar and word formation. To improve transparency, I propose the following steps:

- Romanization of the personal name
- Creation of a latinized stem from the romanized personal name
- Assignment to a declension to create a genitive form suitable as a species epithet
- Addition of suffixes or compound word components to the latinized stem to create genus names or adjectival species epithets.

Romanization of personal names

Most personal names used by authors of scientific publications now originate from writing systems that do not use the Roman alphabet (e.g. Chinese, Japanese, Korean, most Indian languages, Arabic). This means the names have to be romanized, i.e. converted to the modern Roman alphabet. For major world languages, established systems exist for doing this [10]. Where no established romanization scheme exists (e.g. for Georgian, Amharic, Lao, Khmer and Tibetan), names can be transcribed phonetically. However, for any given personal name, there may exist several alternative romanized versions—for example, the most common Chinese surname 王 can be romanized as Wang, Wong, Vang, Ông, Bong, Heng, Vòng, Uōng, Waon or Whang according to the language of origin and romanization scheme used [11]. Similarly, the same romanized form may represent multiple different names in the language of origin; for example the form 'Ji' represents at least 11 different Chinese family names.

As noted in the Botanic Code [12], generally, the romanized form that the author has used in scientific publications is likely to provide the best starting point for the creation of scientific names. However, given that authors may at different times have used different romanizations and choice of a romanization scheme may have personal, cultural or political significance, the author should be asked to confirm how they would like their name to be romanized to create a taxonomic eponym. When an existing romanized component cannot be used to create an agreeable Latin name, alternative options should be explored (see below).

Rule 64 in the ICNP stipulates that diacritic signs are not used in the nomenclature of prokaryotes. This means that where personal names are written with diacritics, the diacritics need to be removed to generate a name consisting of letters from the modern Latin alphabet. In most cases, this simply involves conversion of characters with diacritics to those without (in computational terms, this means conversion from Unicode to ASCII [13]). However, the current version of Rule 64 provides explicit instructions that certain letters with diacritics must be transcribed as digraphs: *ä*, *ö* and *ü* must become *ae*, *oe* and *ue* while *ø*, *æ* and *å* must become *oe*, *ae* and *aa*, respectively. These instructions are culturally insensitive, as the recommended practice for *ä*, *ö* and *ü* is a convention only for names from the German language and does not apply to, for example, Turkish or Finnish names [14]. Thus, the Turkish name Özgür Gündoğdu becomes Ozgur Gundogdu when diacritics are removed, not Oezgur Guendogdu. Generalization from German to other languages has already resulted in formation of the epithet *koenoeneniae* from the Finnish surname Könönen, when kononeniae would be more culturally and aesthetically appropriate. It is thus prudent to stipulate that removal of diacritics should be handled in accordance with customs for the language of origin for the name and with the wishes of the person being honoured.

According to Rule 12a, hyphens need to be removed during romanization. Convention also dictates that all non-alphabetic characters including commas and spaces must be removed, while any internal upper case characters must be converted to lowercase (e.g. in deriving *Yonghaparkia* from Yong-ha Park, *Gaoshiqia* from Gāo Shíqí and *Elizabethkingia* from Elizabeth King).

Creation of a latinized stem

Once rendered in the modern Roman alphabet, a personal name must be converted into a latinized stem, which can be used in word formation. Latinization generally proceeds via incorporation of written characters from the romanized name rather than through an attempt to approximate how the name might be pronounced in Latin. As Pesante points out [15], this leads to the rather unsatisfactory situation in which part of the name (e.g. Escherich in *Escherichia* or Wenzhouxiang *in Wenzhouxiangella*) is usually pronounced as in the source language while remaining morphemes (e.g. suffixes such as *-ia* or *-ella*) are rendered in an approximation of Latin.

However, where a personal name is very long or is judged hard to pronounce in latinized form, there is a long-established custom of simplifying or shortening the personal name during latinization. Examples among names published by Linnaeus include *Clutia*, *Gleditsia*, *Gundelia*, *Mauritia* and *Valantia*—honouring Outger Cluyt, Johann Gottlieb Gleditsch, Andreas von Gundelsheimer, Maurits van Nassau and Sébastien Vaillant, respectively [16–19]. Examples from bacteriology include the genus name *Simkania*, honouring Simona Kahane and the species epithet *chauvoei* named after Auguste Chauveau [6].

Section D(3)(b) in Appendix 9 states that it is not recommended to honour more than one person in one generic name or epithet. However, this practice goes back to the time of Linnaeus, who named the genus *Commelina* after the three Dutch brothers Commelijn and the genus *Bauhinia* after the Bauhin brothers [20]. Furthermore, there are precedents in bacteriology, including *Lechevalieria* after Hubert and Mary Lechevalier [21], *Oceanimonas baumannii* after Paul and Linda Baumann (although here the correct plural form would be *baumanniorum*) [22], and *jeikeium* after Johnson and Kaye. I therefore suggest that this recommendation be removed.

Current guidance in Recommendation 6(7) in the ICNP states that eponyms should be derived only from un-truncated personal names and Section D(3)(c) in Appendix 9 states that if an organism is named after a person, 'the name *cannot* be shortened' (my italics, highlighting the inappropriate modal verb, when this statement is merely advisory). When applied to personal names that are very long and/or hard to pronounce, this guidance can be seen as inconsistent with Recommendations 6(1) and 6(2)—that we should avoid long and disagreeable names—and also flies in the face of the precedents cited above. Furthermore, the advice is inconsistent with Rules 10a and 12c, which state that scientific names can be composed in an arbitrary manner. In fact, as long ago as 1867, de Candolle was arguing that the eponym *Gundelia* was perfectly acceptable, whatever the purists say, if one accepts that it is simply an arbitrary construction [23].

Similarly, when someone has multiple family names—whether linked by white spaces or hyphens—it is generally accepted that it is not mandatory to use a concatenation form of all these names, but a choice should be made as to which components of the personal name best honour the individual while also creating a user-friendly scientific name. Examples here include *Caballeronia* after Jesús Caballero-Mellado, *Franconia* after Augusto Franco-Mora or *Grahamella* after George Stuart Graham-Smith.

For all these reasons, statements ruling out shortening of names should be removed from the guidance and we should allow names to be shortened or simplified as needed, particularly where this helps comply with Recommendations 6(1) and 6(2).

Section D(3)(e) in Appendix 9 states that eponyms can be formed from 'forenames (first names, given names, Christian names), i.e., not from the family name'. Specifying names from a single religion, Christianity, represents one form of unacceptable cultural bias, while use of terms such as 'first name' and 'forename' represents another biased assumption: that all personal names can be shoe-horned into the Anglophone first-name/last-name format. Many, if not most, cultures do not accept this paradigm [24]. As recent publications suggest, when unwilling authors are forced by scientific journals to comply with the first-name/last-name convention, this generates problems, resentment and charges of cultural insensitivity [25–29].

It is worth considering a handful of counter-examples to the first-name/last-name convention. In some Indian and Indonesian cultures, a single personal name is often used. Speakers of Iberian languages tend to use two distinct family names, one inherited from each parent. Some languages adopt naming conventions based on patronyms or matronyms (e.g. Icelandic) or on other family links, e.g. the teknonymic kunyas *Abu* or *Umm* (meaning 'father of' or 'mother of') used by Arabic speakers. Plus, the custom of placing family name before given name (e.g. as in Táng Fēifàn) predominates in East Asia as well as in parts of India and in Hungary.

Section D(3)(e) in Appendix 9 also contains the statement 'First names may be chosen in order to avoid rather long family names or unusually long (hyphenated) double names', while Section D(3)(f) states 'In cases of very frequent family names where the honoured person is not easily identifiable, first and family name may be contracted without connecting vowel or hyphenation'. However, similar problems apply when names are very short (see below). I therefore propose that the recommendations can be usefully simplified to state that any components of a personal name can be used to create a Latin stem for use in the generation of scientific names, so long as this complies with the desire to create names that are as short, agreeable and distinctive as possible. Furthermore, examples in the current advice on the use of given names or given names plus family names are drawn exclusively from European languages. In the interests of inclusivity, examples should also be drawn from other major world languages. Examples here include *Jutongia* after Wú Jùtōng, *Methylobacterium ajmalii* after Ajmal Khan (although stem augmentation is inappropriate here: see below), *Nocardia bhagyanarayanae* after Bhagyanarayana Gaddam, *Yonghaparkia* after Yong-ha Park and *Gaoshiqia* after Gāo Shíqí.

Appendix 9 currently provides some limited advice on how to handle accessory and core components of surnames from a handful of European languages by removing accessory prefixes such as *O'*, *De la* or *van* or combining them with the rest of the name. This approach can be usefully generalized to all languages by giving native speakers the choice of which components they wish to include or omit when latinizing a name.

Stem augmentation: pros and cons

When a latinized stem is created from a modern personal name, the stem is often augmented by addition of an *-i* to the end of the original name. According to authorities such as Nicolson [30] and MacAdoo [31], this practice originates from classical times, when noble Roman families were identified by a *nomen gentilicium* built from the name of an ancestor by addition of an *-i* between the original stem and the declinable ending of the name. For example, the names 'Julius' or 'Julia' were built this way from the ancestral name 'Julus'. Since the renaissance, scholars have often built latinized names using stem augmentation—for example, in latinizing the surname Wezel as *Vesalius*. This practice has been carried over into taxonomic nomenclature when modern surnames are used to create scientific names for genera or species epithets.

However, inconsistencies in the addition of the extra *-i* have plagued taxonomic nomenclature over the centuries. In the early years of modern botanical taxonomy, it was typically applied to species epithets such as *schoepfii* but with exceptions including *richardsoni* and *englemani*. At an early stage in botanical nomenclature, a consensus emerged that the extra *-i* need not be added to surnames ending in *-er*. MacAdoo [31] suggests that this reflects the descriptive rather than hereditary nature of names like *Fisher* or *Hooker*, which means they cannot be treated as *nomina gentilicia*. Nicolson [30] also points out such names also resemble second-declension Latin nouns of the form *puer*, *pueri*, so genitive forms like *fisheri* or *hookeri* were probably thought to sound more natural. Nicolson [30] notes that stem augmentation is generally avoided in botanical nomenclature when the latinized stem ends in a vowel to avoid long strings of vowels, for example in preferring *glazioui* and *glaziouana* to *glaziouii* and *glaziouiana*.

This issue has been treated inconsistently within prokaryotic nomenclature. Thus, we have *Buchnera*, but also *Neisseria* and *Listeria*, *Salmonella* but also *Coxiella* or *Sutcliffiella*. We have the species epithets *youngii* in *Xanthomonas youngii* and *youngiae* in *Desulfosporosinus youngiae* but *youngae* in *Citrobacter youngae*.

The chief argument against adding the additional *-i* is that it creates longer and less euphonious names [32]. Thus, *youngii* is harder to pronounce than *youngi*, while the validly published name *Youngiibacter*—with the inelegant internal 'ii' which is never found inside compound words in classical Latin—is less agreeable and user-friendly than the hypothetical alternative *Youngibacter*.

Although the ICNP does not use the term stem augmentation, current guidance in the ICNP implicitly assumes that stem augmentation is

- essential when forming genitive species epithets or compound words from a personal name that ends in a consonant other than 'r' (in Appendix 9 Tables 2 and 5)
- optional when forming genus names with the suffix *–ella*, even when the names ends in 'r' (in Appendix 9 Table 2)
- excluded when forming adjectival species epithets using the suffixes *-ana*, *-anus*, *-anum* (in Appendix 9 Table 4).

However, current guidance fails to recognize that stem augmentation cannot be justified when applied to components of a name that are not analogous to a *nomen gentilicium*, such as when forming epithets from given names e.g. *jeanii* after Jean Swings or *ruishenii*, after Jiāo Ruìshén.

One complication here is that it is unclear whether creation of a genus name like *Youngia* from the personal name *Young* should be seen as stem augmentation followed by addition of a declinable ending *-a* or whether this is best explained by single-step addition of the suffix *-ia*. This suffix has been used widely in Latin and Greek (and in modern languages) to create abstract nouns, for example in creating names for countries such as *Australia* or medical conditions such as *dementia*. As we have noted, use of this approach to derive plant names from personal names (e.g. *Artemisia*, *Paeonia* and *Euphorbia*) predates the Roman Empire. The Oxford English Dictionary also endorses this interpretation of how such names are formed [33]. If we accept this interpretation, it then makes adding *-ia* to the stem the default procedure, without invoking stem augmentation as an intermediary step.

However, for surnames ending in *-r*, the current advice in the ICNP treats this as a case of stem augmentation, when allowing the option of adding *-a* or *-ia* to create a simple genus name. Scrutiny of precedents within prokaryotic nomenclature reveals that the-–*a* option has been used in only seventeen cases, including *Buchnera* and *Truepera*, while the *-ia* option has been used in 25 cases, including well-known names such as *Burkholderia*, *Listeria* and *Neisseria*.

Given such widespread inconsistency in its use, stem augmentation cannot be considered universally advisable in prokaryotic nomenclature, particularly when it conflicts with Recommendations 6(1) and 6(2) in leading to disagreeable outcomes. In addition, there may be cultural grounds for avoiding this practice. Nicolson [30] argues that stem augmentation honours modern scientists with a privilege afforded originally to patrician Roman families. However, one could argue that adopting naming conventions associated with an ancient imperial social class associated with elitism, patriarchy, corruption, slavery and genocide sits uneasily with modern sensitivities. We can usefully draw on the precedent of zoological nomenclature [34], which avoids the extra *-i* when creating genitive species epithets such as *smithi* or *smithae*. In addition, if Paul Lawson prefers the validly published name *Lawsonibacter* to the hypothetical 'corrected' form *Lawsoniibacter* (Lawson P., personal communication, 2023) or Jonathan Eisen prefers the validly published name *Eisenibacter* to to the hypothetical 'corrected' form *Eiseniibacter* (Eisen J., personal communication, 2023), it seems logical that we should respect that.

However, it is probably not useful to deprecate stem augmentation across the whole of prokaryotic nomenclature. The chief argument for allowing optional stem augmentation is that this broadens the range of possibilities in creating names for prokaryotes, particularly given the need to avoid names already in use in botany or zoology, e.g. in allowing use of *Palleniella* [35] as a bacterial genus name, when *Pallenella* has already been used in zoology. In addition, some microbiologists have become so used to stem augmentation in species epithets that, to their ears, *youngii* might sound more natural than *youngi*.

It would thus seem prudent to advise that stem augmentation is avoided for given names and in situations it creates disagreeable outcomes, e.g. inelegant strings of vowels, but remains allowed in other settings. This would accommodate those who prefer *youngii* to *youngi*, but prevent us from being forced to 'correct' the validly published name *Youngimonas* to *Youngiimonas*. However, when dealing with personal names ending in *-r*, allowing the option of creating simple genus names with *-a* rather than *-ia* just creates unnecessary complexity and inconsistency to no advantage, so I suggest this option is abandoned.

Assignment to a declension

In Latin, unlike in many modern languages, nouns take endings that reflect their role in a sentence—a process known as *inflection*. These grammatical roles are grouped into six so-called *cases*, only two of which are relevant in taxonomic nomenclature:

- the nominative case, which is used for the headword in a dictionary and when the noun is used create a species epithet in so-called *apposition* —for example *gigas*, meaning 'giant' in *Desulfovibrio gigas*
- the genitive form is used to show possession or an association between two nouns and is used in taxonomic nomenclature to create species epithets such as *youngae*.

Latin nouns fall into five categories called *declensions* that govern how inflectional suffixes are applied to the basic form of the word, which the ICNP calls *the stem* (note that this term is used slightly differently in some Latin grammars). Generally, only the first, second and third declensions are used for personal names.

In the first and second declensions, the stem can be identified by removing the inflectional ending from any form of the noun. Thus for the names Julia or Julius, the stem is made by subtracting the inflectional suffix to give us *juli-*. However, in the third declension, the stem typically differs from the form seen in the nominative. So, for the Latin name Cato, the genitive form is *Catonis*, giving us the stem *caton-*, as used in the genus name *Catonella*, honouring Elizabeth Cato.

If the personal name is already a Latin personal name or Latin word, then it is usual to select the latinized stem and declension already associated with the term. The ICNP cites the names *Thomas* and *Cosmas* as examples of first-declension masculine nouns with a stem revealed by removing the final *-as* and a genitive form ending in *-ae*. Two other names in

current use, *Andreas* and *Tobias,* follow the same pattern. The following personal names in current use that represent second-declension Latin names with stems obtained by removing the final -*us* and a genitive form ending in -*i*: Ambrosius, Antonius, Cornelius, Curtius, Erasmus, Flavius, Gallus, Hieronymus, Ignatius, Justus, Laurentius, Marcus, Marius, Paulus, Petrus, Pistorius, Remus and Titus. The name Alexander belongs to the second declension with the stem *alexandr-* with a genitive form ending in -*i*. The contemporary names Michael, Raphael, Ruben and Simon belong to the third declension with stems identical to the nominative form and genitive forms ending in -*is*. The contemporary names Felix, Leo, Paris and Vitalis also belong to the third declension with stems *felic-*, *leon-*, *parid-* and *vital-* and genitive forms ending in -*is*. The name Jesus, often used as a given or family name, is treated as an irregular fourth-declension noun with the stem *jesu-* and the genitive form *jesu*.

It is worth noting that when Latin personal names are used in prokaryotic nomenclature, the stem has not always been selected according to classical usage: for example in deriving the genus name *Rexia* from the personal name Rex (a third-declension noun with the stem *reg-*) or in deriving the species epithet *marcusii* from the surname Marcus (as noted, a second-declension noun with correct genitive form *marci*).

HANDLING OF NAMES ACCORDING TO FINAL LETTER

Handling of names ending in a consonant

To formulate a logical and consistent approach, it is worth reviewing past practice and the current advice enshrined in the Appendix 9 tables. For names ending in a consonant, we can reverse engineer a straightforward workflow from the tables in the ICNP. If a modern name ends in a consonant, then on latinization, if the name bearer is female, the default is to add -*a* (genitive form -*ae*) to create a first-declension feminine noun. If the name bearer is male, then the default is to add -*us* (genitive form -*i*) to create a second-generation masculine noun. In these cases, the stem is made by subtracting the inflectional suffix. Thus, from the modern surname Young, we can derive the declined forms *Younga*, *Youngae* or *Youngus*, *Youngi* together with the stem *young-*. If stem augmentation is applied we also create the declined forms *Youngia*, *Youngiae* or *Youngius, Youngi* and the stem *youngi-*. Then, in accordance with the existing Appendix 9 Table 2, we have the option of creating a genus name from the stem by adding the declinable suffix –*ia* or the diminutive ending -*ella*.

If we wish to form a genus name by deploying a personal name within a compound word, we can follow Appendix 9 Table 3 to join the stem from the personal name with at least one other word component, adding where necessary a connecting vowel. However, as noted above, here it is advisable to suppress stem augmentation, for example, as when the genus name *Youngimonas* was created from the stem Young and the Latin word *monas*. However, when forming names of higher level taxa from the genus name, we should take care to use the appropriate stem, which is derived from the stem of the final word component, which here means we should use *Youngimonad-* as the stem, for example when creating a new family name *Youngimomadaceae*.

Next, we can follow Appendix 9 Table 4 to create an adjectival species epithet by adding the relevant ending according to gender, -*ana*, -*anus* or -*anum*, accepting that stem augmentation here is possible (and preferred by botanists) but brings no benefits. Finally, in line with Appendix 9 Table 5, to create a species epithet in the genitive, we select the appropriate genitive ending: if the holder of the name is female, the default is to select the first-declension female genitive ending -*ae* or if male to add the second-declension masculine ending -*i*. If the name-holder prefers to identify as non-binary, Latin gives us the option of assigning the species epithet to the neuter gender within the second declension, where the inflectional suffix is also -*i*.

Handling of names ending in a vowel: names ending in -*a*

When personal names end in a vowel, current advice in the ICNP practice is less clear and consistent and sometimes inappropriate. Let us start with names ending in -*a*. When used as genitives or with suffixes or in compound words, such names are usually treated as first-declension nouns, where the stem is obtained by removing the final -*a* and the genitive form ends in -*ae*. Thus,

- in Appendix 9 Table 2, we get from Shiga to *Shigella* via the stem *shig-*
- in Appendix 9 Table 4 we similarly get from Loya and Migula to *loyana* and *migulanus* via stems *loy-* and *migul-*
- in Appendix 9 Table 5 we get *victoriae* from Victoria and *voltae* from Volta as species epithets.

However, when the name for a genus is built from a personal name ending in -*a*, the advice is to add the suffix -*ea* to the nominative form of the name to create a genus name, so that, in the examples in Appendix 9 Table 2, we get *Rochalimaea* from Rocha Lima. Looking more widely, there are numerous other eponyms built this way in prokaryotic nomenclature—from *Ameyamaea* to *Vasilyevaea*—so it seems prudent to leave this convention in place. However, framing it as adding -*ea* to a personal name flies in the face of the usual convention that Latin words are built from stems rather than complete words, so it is more consistent to consider the process as adding -*aea* to the latinized stem rather than adding -*ea* to a modern personal name.

In Appendix 9 Table 3, surnames associated with males ending in *-a* such as Ohara and Volta have been unnecessarily converted to masculine second-declension forms, such as *Oharaeus, Oharaus, Oharaius* and *Voltaus*. These in turn are converted to ugly compound forms *Oharaeibacter, Oharaisarcina, Oharaiispirillum* and *Voltaimonas*. However, this stands in the face of many examples from classical and modern times of masculine first-declension names belonging to males but ending in *-a*: from Agrippa to Sulla. The Italian physicist Alessandro Volta was happy to use the first-declension genitive form of his surname *Voltae* on the title page of *De vi attractiva ignis electrici* and as noted this form is also listed in Appendix 9 Table 5. When treated as first-declension nouns, stems are derived from such names by simply removing the final *-a* to give us *ohar-* and *volt-*, which in turn allow us to generate the simpler and more agreeable compound forms *Oharibacter, Oharisarcina, Oharispirillum* and *Voltimonas*.

The treatment in Appendix 9 Table 3 of names associated with females ending in *-a* is also confusing. Two of the names listed, Johanna and Julia, are already Latin names and so need no further latinization. Deriving *Johannicoccus* from Johanna is uncontroversial. However, the table suggests that Julia be latinized as *Juliaea*, which might be suitable as a genus name, but makes no sense when used as stem to create *Juliaeirhabdus*. Given that *Julia* is already a Latin name, as we have already noted, the correct stem is in fact *juli-*. Similarly, without explanation, the name Mateka is latinized as *Matekaia*, which makes no sense as a stem. It is far simpler to treat Mateka as a declinable first-declension noun, which gives us the stem *matek-* and the resulting genus name *Matekibacterium*.

Aside from the simple option of treating names ending in *-a* as first-declension nouns, Appendix 9 Table 5 unhelpfully presents multiple additional options for formation of specific epithets based on precedents such as *oshimai, pamelaeae* and *zhihuaiae*, without providing any justification for such forms. It is hard to see how these examples might illustrate best practice. Although the species epithet *oshimai* has been used to name two species—*Thermus oshimai* and *Caldicoprobacter oshimai*—to create this form, one would have to latinize the personal name Oshima as *Oshimaus*, when it is simpler and more consistent to treat this name as a first-declension noun and generate the genitive form *oshimae,* as has been done in *Picrophilus oshimae*.

Similarly, to get *pamelaeae* from Pamela requires creation of the Latin name *Pamelaea*, which makes sense as a genus name, but not as the basis for a species epithet. Again the simpler and more consistent approach is to treat this name as a first-declension noun and generate the genitive form *pamelae*, as used within the botanical code to name the fungal species *Entoloma pamelae* after its discoverer Pamela Largent.

To get *zhihuaiae* from Zhihua, one would need to latinize the name as *Zhihuaia*, with erroneous application of stem augmentation after a vowel (leading to highly disagreeable five vowels in a row) *and* to a given name (when as noted this approach makes historical sense only for family names). Treating Zhihua as a first-declension noun gives us the far simpler option of *zhihuae*—an approach already used in prokaryotic nomenclature to create the epithet *wanghuae* from the personal name Wang Hua.

Summing up on how to treat personal names ending in *-a*, the simplest and most agreeable option is to treat them all as first-declension nouns, removing the inflexional ending to get the stem and adding *-ae* to derive the genitive form and the suffix *-aea* to generate genus names.

Personal names ending in *-e*

When considering how to treat names ending in *-e*, in Appendix 9 Table 2 the stem of the surname *Benecke* appears to be latinized retaining the final *-e*, to which is added the suffix *-a* to give *Beneckea*, while the suffix *-ia* has been added to the surname *Hoppe* to generate genus name *Hoppeia*. When considering what should be recommended practice here, it is worth looking at a range of precedents. Addition of the suffix *-a* to personal names ending in *-e* to form genus names is the more commonly used option, with validly published examples among prokaryotes including *Bosea, Coetzeea, Fannyhessea, Hoeflea, Lentzea, Levinea, Lonsdalea, Savagea, Sharpea* and *Xaviernesmea*. This is also recommended practice in botanical nomenclature. By contrast, aside from *Hoppeia*, addition of *-ia* is seen only in *Agreia, Leeia, Mooreia* and *Woeseia*. It would thus seem prudent to recommend the shorter and simpler option, which also avoids the need for three or more vowels in a row.

The addition in Appendix 9 Table 2 of the diminutive suffix *-lla* instead of *-ella* to create *Brucella* from Bruce makes sense as it avoids two copies of the letter *-e* in a row. However, when this approach is applied to personal names ending in the double *-e* digraph, it can generate names with uncertain pronunciation, such as *Crabtreella* and *Hyunsoonleella*. When it occurs at the ends of names, this digraph is probably best latinized as the letter, e.g. *Macgiella* rather than *Macgeella* from McGee. In Appendix 9 Tables 3–5, the final *-e* has been retained during latinization of the surnames Pace, Curie, Love and Beveridge to generate *pacei, curieae, loveana* and *beveridgei* as epithets.

Summing up on how to treat personal names ending in *-e*, it seems advisable to recommend

- retaining the final *-e* when generating a latinized stem

- forming genus names from the stem using the suffix *-a* rather than *-ia*
- eliding the *-e* from the diminutive ending, using *-lla* rather than *-ella*
- to latinize the digraph '*ee*' as '*i*' at the end of personal names, particularly when forming diminutives.

Personal names ending in *-i* and *-y*

In Appendix 9 Table 2, for surnames ending in *-i*, the original final *-i* is left as part of the latinized stem, whether one is creating a genus name with the simple suffix *-a* or the diminutive *-ella*, to give us *Nevskia* and *Terazakiella*. Such names and stems are treated similarly in Appendix 9 Tables 3–5. However, in compound names given as examples, a cumbersome connecting vowel has been added to create the disagreeable forms *Terasakiispira* and *Yabuuchiispira*. Continuing recommendations to treat the final *-i* as part of the stem seems uncontroversial. However, drawing on the example set in Section G(2) of Appendix 9, which states 'in view of the many precedents in the past, addition of a connecting vowel after *alkali-* is not recommended', I propose that addition of a connecting vowel after a stem ending in *-i* is also not recommended when creating compound nouns from personal names.

Names ending in *-y* are treated similarly, although addition of a connecting vowel in compound names here is less problematic. In line with botanical nomenclature [30], changing a *-y* to an *-i* can create an agreeable form more in line with classical Latin usage—as in deriving the bacterial genus name *Cowdria* and *Roseburia* from the personal names Cowdry and Rosebury or the species epithet *asburiae* after the surname Fife-Asbury.

Personal names ending in *-o*

In Appendix 9 Tables 2 and 4, the surnames Beggiato and Guerrero are latinized with unchanged stems to give us *Beggiatoa* and *guerreroanus*. However, In Appendix 9 Tables 2 and 4, such names are more commonly treated as third-declension nouns with the genitive form *-onis*, which gives us a stem ending with *-on* to which relevant suffixes are added—so we have *Seinonella* from Seino, *hiranonis* from Hirano and *Kozakonia* from Kozako. In Appendix 9 Table 3, the surname Franco is also treated as a third-declension noun with the stem *francon-*, although unnecessary complexity is added by offering an additional option for latinization as the second-declension stem-augmented *Franconius*. Similarly, in Appendix 9 Tables 3 and 5, it is suggested that the name *Cato* should be latinized as the first-declension stem-augmented *Catonia*, when Cato is already a Latin personal name with the stem *caton-* and the genitive form *catonis*.

As before, when considering what should be recommended practice here, it is worth looking at a range of precedents. In prokaryotic nomenclature, personal names ending in *-o* have most commonly been treated as third-declension nouns. Examples of simple genus names include *Bizionia*, *Caballeronia*, *Geojedonia*, *Hwangdonia*, *Isachenkonia*, *Jejudonia*, *Kosakonia*, *Nesterenkonia*, *Saonia*, *Wandonia* and *Zarconia*, while those formed with the diminutive include *Bailinhaonella*, *Lutaonella*, *Seinonella*, *Taonella* and *Zhaonella*. Examples of genitive species epithets formed in this way include *bennonis*, *caballeronis*, *caoxuetaonis*, *claudionis*, *distasonis*, *endonis*, *franconis*, *ghigonis*, *hatanonis*, *hiranonis*, *hongtaonis*, *indalonis*, *kanbiaonis*, *kanungonis*, *ruizarguesonis*, *salanitronis*, *salomonis*, *subbaraonis*, *togonis*, *tojonis*, *uenonis*, *xianingshaonis* and *xujianguonis*. In contrast, aside from *Beggiatoa*, I can find only a handful of prokaryotic names have avoided adding the letter 'n' to the stem: *Asanoa*, *Ligaoa*, *Montanoa* and *miyamotoi*. Here, custom and practice in prokaryotic nomenclature has diverged from botanic nomenclature, where treating modern names as third-declension nouns is deprecated [30].

There is an added complication in that, in some cases, a third-declension stem ending in 'n' has been created, but then subjected to stem augmentation and shifted to the first or second declension. Thus, in addition to *catoniae*, we also have first-declension epithets such as *araujoniae*, *albertanoniae*, *trujilloniae*, *toranzoniae* and *ndongoniae* and second-declension epithets such as *asenjonii*, *bizzozeronii*, *branconii*, *idonii*, *ishimotonii*, *issachenkonii*, *shilonii* and *qiguonii*. Here, an apparent desire to mark the gender of the name bearer ignores the recommendations to keep names as short and simple as possible.

In conclusion, I suggest that in prokaryotic nomenclature, we restrict ourselves to the simplest option in recommending that names ending in *-o* should be treated as third-declension nouns with the genitive form *-onis* and a stem ending in *-on*.

Personal names ending in *-u*

From the examples in Appendix 9, when personal surnames end in *-u*, the final *-u* is left as part of the latinized stem, but an extra *-i* is recommended when creating simple genus names and species epithets in the genitive. Therefore, in these examples, we have *Simiduia* from *Simidu*, *quenuiae* from Plateau-Quénu, *brisouii* from Brisou. Similarly, an extra *-i* is added to when creating the fictional compound name *Letestuiinema* from the family name Le Testu. However, stem augmentation has not been applied when creating the fictional example *manescuanus* from Manescu, nor in creating *Shimazuella* from Shimazu.

Consideration of a larger set of examples suggests that the suffix *-ia* has generally been used to create genus names from personal names ending in *-u*, with examples such as *Jejuia*, *Minwuia*, *Puia*, *Xuhuaishuia* and *Zhouia*, while stem augmentation has been

applied to create genitive epithets such as *gauvreauii*, *neuii*, *simiduii*, *qiuiae* and *zeshuii*. The extra -*i* has not so far been used in diminutives derived from names ending in -*u* but does appear when deriving *wuianus* from the name Wu.

In line with botanical nomenclature, the simpler and more consistent option would be to avoid adding an extra 'i' whenever a name ends in a vowel, which means simply adding -*a* to stems ending in -*u* to create genus names such as *Letestua* and would avoid long strings of vowels in species epithets. This would make practice with names ending in -*u* consistent with recommendations for names ending in other vowels.

ANALYSIS OF A GLOBAL DATASET

PubMed last names

Precedents from prokaryotic nomenclature can only take us so far, particularly when a strong cultural bias towards the Western world remains among prokaryotic eponyms. To determine how the recommendations in the ICNP might play out with a representative global set of personal names used by contemporary scientists, five consecutive daily XML files from August 2023—pubmed23n1431.xml, pubmed23n1432.xml, pubmed23n1433.xml, pubmed23n1434.xml and pubmed23n1435.xml—were downloaded from PubMed on 17 August 2023 from the site: https://ftp.ncbi.nlm.nih.gov/pubmed/updatefiles/ and decompressed. The Python script *pubmed_surname_parser.py* (Accessory File 1) [1] was used to extract the contents of the LastName field of all authors, which on de-replication gave ~125000 unique last names, which the script sorted by frequency, length and alphabetically (Supplementary Data Sheet 1, Accessory File 2) [1].

Note that the LastName field is derived from article metadata provided by publishers when articles are indexed for the PubMed database. Publishers typically follow the standard practice of using the names as they appear in the published article, which is determined by the authors themselves prior to submission. However, text not representing last names will find its way into this field if errors or divergent practices occur at data entry (e.g. when other components of personal names are included, when authors are forced to add text to the field even though they do not have a last name or when names of research consortia are added).

Scrutiny of names by frequency reveals that the commonest last names in publications listed in PubMed predominantly reflect East Asian or South Asian heritage. The only last names of unequivocally European origin in the top 100 are Smith, Jones, Williams, Brown, Johnson, Thomas, Taylor, Wilson and Martin. Scrutiny of names by length reveals nineteen last names consisting of a single letter (A, B, C, D, E, F, G, H, I, J, K, M, O, P, R, S, T, U, V) while over 150 consist of just two letters, including 15 of the top 100 names. There are seven last names without hyphens or spaces with 20 or more letters: Chalermchockcharoenkit, Chaithongwongwatthana, Watanaboonyongcharoen, Detsakunathiwatchara, Khongphatthanayothin, Randrianomenjanahary and Sirimongkolchaiyakul. The longest hyphenated last name is Martín-Fernández-de-Labastida, with 29 characters. The longest entry in a LastName field with spaces is Ignacio Morgado García de Polavieja, with 35 characters (although seems likely that Ignacio here is actually part of the forename). At least 19 last names consist only of vowels (A, Ae, Ai, Ao, Aoi, Au, Aue, E, Eo, I, Iio, O, Oe, Oei, Oi, Oo, Ooe, Ooi, Ou), while there are over 40 last names without any vowels, including two that occur in the top two hundred names: Ng and Lv.

Fine-tuning of current recommendations

Although some of the entries in the LastName field clearly do not reflect personal names (e.g. *Intelligence And Neuroscience*) and some might represent errors in entering data, checking some of the shortest and longest names via searches on the PubMed website (e.g. https://pubmed.ncbi.nlm.nih.gov/?term=A[Au] and https://pubmed.ncbi.nlm.nih.gov/?term=Mart%C3%ADn-Fern%C3%A1ndez-de-Labastida[Au]) confirms that they generally do represent genuine personal names and thus shows that current recommendations in Appendix 9 are incomplete and even unhelpful in dealing with the full variation seen in last names. Therefore, for example, when faced with family names such as Chalermchockcharoenkit, Chaithongwongwatthana or Watanaboonyongcharoen, it makes no sense to insist that a family name can only be used in its entirety. Instead, in line with the precedents already cited (*Gundelia*, *Simkania*, *etc.*), it makes sense to truncate stems to create short agreeable names such as *Chalermia*, *Chaithongia* or *Watania*.

Similarly, a family name consisting of just one or two letters generally cannot be used it to create agreeable and distinctive scientific names if we follow current advice. For example, if we uncritically follow the advice in the ICNP when handling the family names A and Ma—by dropping the final -*a* before adding a suffix—we end up with the genus names *Ea*, *Ella*, *Mea* and *Mella* and the species epithets *ae*, *ana*, *anus*, *anum* and *mae*, *mana*, *manus*, *manum*. In such cases, either of the two options—use of a given name or given name and family name combined—usefully apply. Thus to honour a microbiologist called Mǎ Wénbó, one could create *Mawenbonella* or *Wenbonella*, while honouring a scientist called Runa A, one could use *runana* or *arunana* as an adjectival species epithet.

Similar problems apply when romanization generates forms that lack vowels. Examples here include the Chinese surname Lv (an attempt to write the Hanyun Pinyin version of the name, Lü, without diacritics) and the Cantonese surname Ng. These names can be alternatively romanized as *Lyu*, *Lu* or *Lui* and *Ang*, *Eng*, *Ing*, *Ong* and *Ung* to create agreeable scientific names such as

Lyuella or *Ungia*. When faced with long strings of consonants or vowels, just as in Linnaeus' time, it remains advisable to create a simplified form that better replicates Latin phonology—for example we can derive the theoretical names *Henslegeria* from Hengstschläger or *Pribilskia* from Przybylski.

Automation of name creation

Drawing on the above review and analysis, I have devised a simple unified workflow (which is presented below as Table 1 in the proposed emendation). To enable computer-based generation of names using this workflow, I then created a Python script (Accessory File 3) [1] that takes as input a text file listing personal names and generates as output a list of genus names and species epithets built from these personal names.

To create an extensive dataset of eponyms for use in prokaryotic nomenclature, the Python script was applied to the de-replicated list of ~125000 last names extracted from PubMed to generate around 1.6 million genus names and over 1 million species epithets (Accessory Files 4 and 5) [1]. The script incorporated a set of steps needed to comply with relevant advice:

- If the personal name was longer than 18 characters, name creation was suppressed and a comment was added: 'This name is too long to create a user-friendly taxonomic name. You should consider using a shortened form or another component from the personal name'.
- If the personal name was shorter than three characters, name creation was suppressed and a comment was added: 'This name is probably too short to create a distinctive and user-friendly taxonomic name. You should consider using additional components of the name'.
- If the personal name contained no vowels, name creation was suppressed and a comment was added: 'This name contains no vowels. Please use an alternative approach to romanizing the name that adds some vowels'.
- The characters *ø*, *æ* and *å* were replaced with *oe*, *ae* and *aa*. If the personal name was found to originate in the German language, *ä* was replaced by *ae*, *ö* by *oe* and *ü* by *ue*. In all other cases, characters with diacritics were replaced by the same characters without diacritics. In all such cases, the comment was added: 'This name contains letters from outside the simple Roman alphabet. Please check that these have been romanized in accordance with customs for the language of origin'.

As noted above, some of the entries in the LastName field may not reflect genuine last names but nonetheless the script will generate genus names and species epithets from these entries. However, manual curation of a dataset of this size is not possible, so the rare cases with erroneous outputs should be ignored.

CONCLUSIONS

Here, I have shown that current recommendations within the ICNP governing creation of prokaryotic eponyms require updating to become less culturally biased, easier to follow and more consistent. I have made available a large set of eponyms built from last names in the PubMed database. A key limitation is the current approach does not yet enable automated creation of scientific names from other components of personal names nor does it check whether the name has already been used in taxonomy. However, the principles articulated here will facilitate customized name creation on a case-by-case basis and lay the groundwork for more sophisticated approaches to automation.

PROPOSED EMENDATIONS TO THE ICNP

On the basis of the above review and analyses, I propose the following changes to the ICNP.

The following text

RECOMMENDATION 6

(7) If genus names or specific epithets are formed from personal names, they should contain only the untruncated family (rarely given) name of a person. Authors should not name organisms after themselves or co-authors.

Should be changed to:

RECOMMENDATION 6

(7) Authors should not name organisms after themselves or co-authors.

The following text

Diacritic Signs
Rule 64
Diacritic signs are not used in the nomenclature of prokaryotes.

In names or epithets derived from words with such signs, the signs must be suppressed and the letters transcribed as follows: (1) *ä*, *ö* and *ü* become *ae*, *oe* and *ue*; (2) *é*, *è* and *ê* become *e*; (3) *ø*, *æ* and *å* become *oe*, *ae* and *aa*, respectively.

Should be changed to:

Diacritic Signs
Rule 64
Diacritic signs are not used in the nomenclature of prokaryotes. In names or epithets derived from words with such signs, the signs must be suppressed and the letters transcribed in accordance with established customs for their language of origin.

To facilitate reading and due to its length, the current text of Appendix 9 Section D is not reproduced here but can be found in pp 74–76 of the ICNP [6]. The body text of Appendix 9 Section D should be replaced with the following text, while Tables 2, 3, 4 and 5 should be replaced with Table 1 below.

D. Formation of prokaryote names from personal names

(1) Persons may be honoured by using their name in forming a generic name or a specific epithet. However, the *Code* recommends refraining from naming taxa after persons that are not connected with microbiology or, at least, with natural science. It is good practice, where possible, to ask the person to be honoured by a scientific name for permission to use their name and to take into account their preferences for how the name is latinized.

Authors should refrain from naming bacteria after themselves or co-authors in the same publication [see Recommendations 6(7) and 12 c(3)]. The formation of prokaryote names from personal names has no geopolitical meaning, i.e., the way in which such names are formed cannot be used to express geopolitical claims [see General Consideration 8]. More than one person can be honoured in a single generic name or epithet. Examples: *Lechevalieria* after Hubert and Mary Lechevalier and *Preeria* after Louise and John Preer.

(2) Latinization of personal names

(a) When the personal name is already a Latin name, no further latiniszation is needed, e.g. Julia, Victoria, Leo. When originating from languages written in non-Roman scripts, personal names should be rendered in the Roman alphabet according to the existing romanization scheme(s) for the name-bearer's language or through transcription of the spoken form where no such scheme exists.

Where there is a choice as to how a name should be romanized, the wishes of the person to be honoured should be taken into account. Hyphens and white spaces should be removed from the name [see Rule 12a]. Diacritics should be removed according to the conventions of the language of the name-bearer (e.g. German surnames Höffner and Müller become Hoeffner and Mueller, whereas the Turkish surname Özgür becomes Ozgur and the Finnish surname Törönen becomes Toronen) [See Rule 64].

(b) A romanized derivative of the personal name should be used to create a latinized stem to which suffixes can be added or which can be used in compound word formation. Choice of which parts of the personal name are used to create the latinized stem should take into account the wishes of the person being honoured, relevant traditions (e.g. in cultures where multiple family names are used or none at all) and the desire to create agreeable user-friendly scientific names. For example, when someone has multiple family names, just one of these can be used to honour them. Example: *Boudabousia*, honouring Abdellatif Boudabous Chihi, *Streptomyces lunalinharesii* honouring Luiz Fernando de Toledo Luna Linhares, *Citrobacter murliniae* honouring Alma C. McWhorter-Murlin.

Although a family name is most often used, generic names or specific epithets can also be formed from given names. Examples: *Erwinia* after Erwin Smith, *Jutongia* after Wú Jùtōng, *Staphylococcus arlettae* after Arlette van de Kerckhove and *Nocardia bhagyanarayanae* honouring Bhagyanarayana Gaddam. Scientific names can also be formed from given names placed before or after family names conjoined without a connecting vowel. Examples: *Elizabethkingia* honouring Elizabeth King, *Yonghaparkia* honouring Yong-ha Park and *Gaoshiqia* honouring Gāo Shíqí.

Kinship prefixes (e.g. 'O'/'Ni', 'Mac/Mc', 'ibn/bin/bint', 'Abu', 'ap' and 'Ben') and particles and articles (e.g. 'al', 'de', 'le', 'van' and 'von') may be omitted or included in the latinized stem. Examples: *Rochalimaea* after da Rocha-Lima and *Leclercia* after Le Clerc. However, honorific titles (e.g. 'Sir', 'Sri' and 'Sheikh') are not usually included in prokaryote names.

To render the latinized stems as short and easy to pronounce as possible, simplified or truncated forms of personal names may be used. Examples *chauvoei* after Chauveau; *Simkania* after Simona Kahane, *jeikeium* after Johnson and Kaye) and

additional vowels may be added (e.g. *macginleyi* after Kenneth John McGinley). Intentional latinizations involving changes in orthography of personal names must be preserved [see Rule 60].

When the personal name ends in -*a*, the name should be treated as a first-declension Latin noun, where the stem is formed by removing the -*a*. When the personal name is a Latin name that ends in -*us*, the name should be treated as a second-declension Latin noun, where the stem is formed by removing the -*us*. When the personal name ends in -*o*, the name should be treated as a third-declension Latin noun, where the stem is formed by adding 'n' to give a genitive form ending in -*onis*.

When a family name is used, there is the option to augment the stem by adding an -*i* to the end of the name, e.g. as in the epithet *youngii* in *Xanthomonas youngii*, after John Young. However, this need not happen, as seen in *Citrobacter youngae*, which honours Viola Young. Stem augmentation is generally avoided when the latinized stem ends in a vowel, e.g. *voltae* not *voltiae* from Volta; *sakazakii* not *sakazakiii* from Sakazaki. Some authorities suggest stem augmentation should avoided for names ending in -*er*, e.g. in *Buchnera*. However, for consistency, it is advisable to add the suffix -*ia* when forming genus names in such cases e.g. *Listeria* or *Burkholderia*.

3) *Personal names in generic names*

There are three suggested ways to form a generic name from the latinized stem of a personal name

(i) directly, by adding the feminine ending -*ia* (or -*a* if the latinized stem ends in *e*, *i* or *u*, or -*aea* when the original personal name ends in -*a*)

(ii) as a diminutive, usually by adding the feminine ending -*ella* to the stem. When the stem ends in -*e*, the final letter should be omitted when forming diminutives, e.g. in forming *Brucella* from the personal name Bruce. When digraph *'ee'* as occurs at the end of personal names it should be latinised as *'i'* when forming diminutives, e.g. the hypothetical name *Macgiella* is preferred to *Macgeella* when forming a name after McGee.

(iii) by using a stem derived from a personal name as a word element in a compound name. Here the stem is linked, as needed, to the following word element by a connecting vowel, e.g. *Youngimonas*, derived from the surname *Young* and N.L. fem. n. *monas*. When the latiniz*ed stem ends in -*i, the connecting vowel is omitted.*

(4) Personal names in specific epithets

Two possibilities exist to form specific epithets from personal names

(i) creation of a genitive noun by addition of an inflectional ending to the latinized stem so that an epithet is formed with the meaning of 'pertaining to the person being honoured'. The inflectional ending is chosen to reflect the sex of the person to be honoured: typically the feminine ending -*ae* for a female or the masculine or neuter ending -*i* for a male or non-binary individual, except where a male name belongs to first declension, e.g. *voltae* from Volta or a name from a female belongs to the second declension, e.g. *pistorii* to honour a female with the surname *Pistorius* or where a name of any gender has been assigned to the third declension, e.g. *mayonis* after Mark Mayo. Where the personal name resembles a Latin genitive it may be used unchanged as a species epithet, e.g. *imshenetskii* to honour Aleksandr Imshenetskii

(ii) creation of an adjectival form by adding the endings -*anus* (masculine), -*ana* (feminine) or -*anum* (neuter) to the stem according to gender of the genus name. Care should be taken to avoid creation of epithets with unwanted meanings, e.g. *loveanus* from the surname Love.

TABLE 1 UNIFIED WORKFLOW FOR CREATION OF PROKARYOTIC EPONYMS FROM PERSONAL NAMES

Romanization of the personal name

- If the personal name originates from a language not written in the Roman alphabet, romanize the personal name in accordance with conventions for the language of origin and in accordance with the preferences of the personal being honoured. When no suitable romanization system exists, transcribe the name phonetically.

Creation of a latinized stem

- Choose one or more components of the personal name suitable for creation of an agreeable and distinctive scientific name. Avoid components that are too short, too long or otherwise difficult to render in Latin (e.g. containing too few or too many vowels or consonants). Where a single component of the name cannot be on it own used to create an agreeable and distinctive latinized stem, use additional or alternative components of the personal name (e.g. a given name or a given name combined with a family name) or explore additional options for romanization.
- Remove diacritics, punctuation marks, hyphens and spaces to create a latinized stem. Replacement of characters with diacritics should be undertaken in accordance with customs for the language of origin. If the personal name originates in the German language, characters with umlauts should be replaced by an appropriate digraph: *ä* is replaced by *ae*, *ö*

by *oe* and *ü* by *ue*. However, in other languages using such characters (e.g. Turkish or Finnish) the vowel with an umlaut should be replaced by the same vowel without the umlaut. For personal names originating from Nordic languages, *ø*, *æ* and *å* become *oe*, *ae* and *aa*.

- For names of Gaelic origin, expand the patronymic prefix *Mc* to *Mac*.
- Convert the stem to lower case.
- Optional: if the personal name ends in a consonant add an *-i* to the original stem to create an augmented stem.

Assignment to a declension

- If the personal name is already a Latin personal name or Latin word, then select the latinized stem and declension already associated with the term (e.g. first declension if name ends in *-a*, second declension if the name ends in *-us*). The names *Andreas*, *Cosmas*, *Thomas* and *Tobias* are assigned to the first declension with a stem revealed by removing the final 'as'. *Alexander* is assigned to the second declension with the stem *alexandr-*. The names *Michael*, *Raphael*, *Ruben* and *Simon* are assigned to the third declension with stems identical to the nominative form. The names *Felix*, *Leo*, *Paris* and *Vitalis* are assigned to the third declension with stems *felic-*, *leon-*, *parid-* and *vital-*. The name *Jesus* is treated as an irregular fourth-declension noun with the stem *jesu-* and the genitive form *jesu*.
- If the personal name ends in *-a*, assign the latinized stem to the first declension.
- If the personal name ends in *-o*, add 'n' to end of the name to create a latinized stem and assign the latinized stem to the third declension.
- Otherwise, assign the stem to the first declension if the name bearer is female or to the second declension if the name bearer is male.

To generate a simple genus name

- If the stem belongs to the first declension, add 'aea' to the stem
- If the stem does not belong to the first declension and ends in a consonant, add 'ia' to the stem
- If the stem does not belong to the first declension and ends in a vowel, add *-a* to the stem.
- Convert the first letter of the stem to upper case.
- Convert the genus name to italics.

To generate a diminutive genus name

- If the stem ends in *-e* add '*lla*'
- Otherwise, add '*ella*'
- Convert the first letter of the genus name to upper case
- Convert the genus name to italics.

To generate a compound genus name

- If the stem ends in *-i*, combine the stem and a subsequent word component without a connecting vowel.
- Otherwise, combine the stem and a subsequent word component with the connecting vowel *-i*.
- Convert the first letter of the genus name to upper case.
- Convert the genus name to italics.

To generate an adjectival species epithet

- If the genus name is feminine, add *-ana* to the stem.
- If the genus name is masculine, add *-anus* to the stem.
- If the genus name is neuter, add *-anum* to the stem.
- Convert the species epithet to italics.

To generate a genitive species epithet

- If the stem belongs to the first declension, add 'ae' to the stem.
- If the stem belongs to the second declension, add *-i* to the stem.
- If the stem belongs to the third declension, add 'is' to the stem.

Conflicts of interest
The author declares that there are no conflicts of interest.

References
1. **Pallen M**. Dataset; 2023. https://figshare.com/s/aa95daebca9be767b6b2

2. **Pliny the Elder**. The Natural History (English translation by Bostock and Riley); 1855. http://www.perseus.tufts.edu/hopper/text?doc=Perseus:text:1999.02.0137 [accessed 14 July 2023].

3. **Stearn WT**. *Botanical Latin*. Timber Press (OR), 1995.

4. **Wright J**. *The Naming of the Shrew*. Bloomsbury Publishing, 2015.

5. **Oren A** *et al.* How to name new taxa of Archaea and Bacteria. In: **Whitman WB, Rainey F** and **Kämpfer P** (eds). *Bergey's Manual of Systematics of Archaea and Bacteria*. Chichester, UK, 2015. pp. 1–24.

6. **Oren A, Arahal DR, Göker M, Moore ERB, Rossello-Mora R,** *et al.* International Code of Nomenclature of Prokaryotes. Prokaryotic Code (2022 Revision). *Int J Syst Evol Microbiol* 2023;73.

7. **Pethiyagoda R**. Policing the scientific lexicon: the new colonialism? *MT* 2023;10:20–25.

8. **Pallen MJ, Telatin A, Oren A**. The next million names for Archaea and Bacteria. *Trends Microbiol* 2021;29:289–298.

9. **Pallen MJ, Rodriguez-R LM, Alikhan NF**. Naming the unnamed: over 65,000 *Candidatus* names for unnamed Archaea and Bacteria in the Genome Taxonomy Database. *Int J Syst Evol Microbiol* 2022;72.

10. **UK Government**. Romanization systems[internet]; 2014. https://www.gov.uk/government/publications/romanization-systems [accessed 14 July 2023].

11. **Wikipedia contributors**. Romanization; 2023. https://en.wikipedia.org/wiki/Romanization [accessed 14 July 2023].

12. **Turland NJ, Wiersema JH, Barrie FR,** *et al.* *International Code of Nomenclature for Algae, Fungi, and Plants (Shenzhen Code) Adopted by the Nineteenth International Botanical Congress Shenzhen, China, July 2017*. Koeltz botanical books, 2018.

13. **Burke SM**. Unidecode. *Sys Admin* 2001;10:54–60.

14. **Alpkoçak A, Ceylan M**. Effects of diacritics on Turkish information retrieval. *Turk J Electr Eng Comput Sci* 2012;20:787–804.

15. **Pesante A**. About the use of personal names in taxonomical nomenclature. *Taxon* 1961;10:214–221.

16. **Wikipedia contributors**. List of plant genera named for people (A–C); 2023. https://en.wikipedia.org/wiki/List_of_plant_genera_named_for_people_(A–C) [accessed 14 July 2023].

17. **Wikipedia contributors**. List of plant genera named for people (D-J); 2023. https://en.wikipedia.org/wiki/List_of_plant_genera_named_for_people_(D-J) [accessed 14 July 2023].

18. **Wikipedia contributors**. List of plant genera named for people (K-P); 2023. https://en.wikipedia.org/wiki/List_of_plant_genera_named_for_people_(K–P) [accessed 14 July 2023].

19. **Wikipedia contributors**. List of plant genera named for people (Q-Z); (n.d.). https://en.wikipedia.org/wiki/List_of_plant_genera_named_for_people_(Q–Z) [accessed 14 July 2023].

20. **Austin D**. The nuance and wit of carolus linnaeus. *The Palmetto* 1993;13:8.

21. **Labeda DP, Hatano K, Kroppenstedt RM, Tamura T**. Revival of the genus *Lentzea* and proposal for *Lechevalieria* gen. nov. *Int J Syst Evol Microbiol* 2001;51:1045–1050.

22. **Brown GR, Sutcliffe IC, Cummings SP**. Reclassification of [*Pseudomonas*] doudoroffii (Baumann et al. 1983) into the genus *Oceanomonas* gen. nov. as *Oceanomonas doudoroffii* comb. nov., and description of a phenol-degrading bacterium from estuarine water as *Oceanomonas baumannii* sp. nov. *Int J Syst Evol Microbiol* 2001;51:67–72.

23. **Candolle A de, International Botanical Congress**. *Lois de la nomenclature botanique /*. In: *Masson et Fils*. Paris, 1867.

24. **Wikipedia contributors**. Personal name; 2023. https://en.wikipedia.org/wiki/Personal_name [accessed 14 July 2023].

25. **Gasparyan AY, Yessirkepov M, Gerasimov AN, Kostyukova EI, Kitas GD**. Scientific author names: errors, corrections, and identity profiles. *Biochem Med* 2016;26:169–173.

26. **Raveenthiran V**. Insensitivity of editors and indexers regarding the cultural variations of authors' surnames. *Biochem Med* 2016;26:164–168.

27. **Qiu J**. Scientific publishing: identity crisis. *Nature* 2008;451:766–767.

28. **Sheherazade, Ardiantiono** Attention science: some people have only one name. *Nature* 2020.

29. **Puniamoorthy N, Jeevanandam J, Narayanan Kutty S**. Give south Indian authors their true names. *Nature* 2008;452:530.

30. **Nicolson DH**. Orthography of names and epithets: latinization of personal names. *Taxon* 1974;23:549–561.

31. **MacAdoo TO**. Nomenclatural literacy. In: Goodfellow M, O'Donnell AG and Sneath PHA (eds). *Handbook of New Bacterial Systematics*. San Diego, CA: Academic Press; 1993. pp. 339–358.

32. **Pallen MJ**. Bacterial nomenclature in the era of genomics. *New Microbes New Infect* 2021;44:100942.

33. **Oxford English Dictionary**. "-ia, suffix1". OED Online; 2023. https://www.oed.com/view/Entry/90676?result=2&rskey=g0wfbu& [accessed 14 July 2023].

34. **International_Commission_on_Zoological_Nomenclature**. *International Code of Zoological Nomenclature. Fourth Edition*. International Trust for Zoological Nomenclature, 1999.

35. **Hitch TCA, Bisdorf K, Afrizal A, Riedel T, Overmann J,** *et al.* A taxonomic note on the genus *Prevotella*: description of four novel genera and emended description of the genera *Hallella* and *Xylanibacter*. *Syst Appl Microbiol* 2022;45:126354.