Journal of Biomedical Informatics 55 (2015) 55–63

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

MorphoCol: An ontology-based knowledgebase for the characterisation of clinically significant bacterial colony morphologies

Ana Margarida Sousa [a], Maria Olívia Pereira [a], Anália Lourenço [a,b,*]

[a] CEB – Centre of Biological Engineering, LIBRO – Laboratório de Investigação em Biofilmes Rosário Oliveira, University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal
[b] ESEI: Escuela Superior de Ingeniería Informática, University of Vigo, Edificio Politécnico, Campus Universitario As Lagoas s/n, 32004 Ourense, Spain

ARTICLE INFO

Article history:
Received 29 April 2014
24 January 2015
Accepted 20 March 2015
Available online 25 March 2015

Keywords:
Colony morphology
Antimicrobial resistance
Virulence
Colony morphology ontology

ABSTRACT

Background: One of the major concerns of the biomedical community is the increasing prevalence of antimicrobial resistant microorganisms. Recent findings show that the diversification of colony morphology may be indicative of the expression of virulence factors and increased resistance to antibiotic therapeutics. To transform these findings, and upcoming results, into a valuable clinical decision making tool, colony morphology characterisation should be standardised. Notably, it is important to establish the minimum experimental information necessary to contextualise the environment that originated the colony morphology, and describe the main morphological features associated unambiguously.
Results: This paper presents MorphoCol, a new ontology-based tool for the standardised, consistent and machine-interpretable description of the morphology of colonies formed by human pathogenic bacteria. The Colony Morphology Ontology (CMO) is the first controlled vocabulary addressing the specificities of the morphology of clinically significant bacteria, whereas the MorphoCol publicly Web-accessible knowledgebase is an end-user means to search and compare CMO annotated colony morphotypes. Its ultimate aim is to help correlate the morphological alterations manifested by colony-forming bacteria during infection with their response to the antimicrobial treatments administered.
Conclusions: MorphoCol is the first tool to address bacterial colony morphotyping systematically and deliver a free of charge resource to the community. Hopefully, it may introduce interesting features of analysis on pathogenic behaviour and play a significant role in clinical decision making.
Database URL: http://morphocol.org.

1. Introduction

Human infections involve a complex intertwined interplay of microorganisms. Understanding these interactions as well as the continuously emerging mechanisms of antimicrobial resistance are pressing goals in clinical microbiology.

Recent technologies, such as the enzyme-linked immunosorbent assay (ELISA), the polymerase chain reaction (PCR) and the matrix-assisted laser desorption/ionisation time of flight mass spectrometry (MALDI-TOF MS) have boosted the identification and characterisation of clinically significant bacteria [1–3]. However, the research community is manifesting a renewed interest in traditional culture-based strategies like colony morphology characterisation as more immediate, first-term means of decision support [4,5].

Alterations in the morphology of the microbial colonies, reflected in macroscopically observable features such as form, colour, opacity, size and texture, may support bacteria profiling under changing and often stressful environments [6–8]. Notably, these morphological features are being increasingly documented in clinical settings as potential evidences of the expression of virulence factors [9–12] and increased resistance to antibiotic therapeutics [13–15]. For example, mucoid morphotypes [16,17] and small colony variants [18–20] are recognised as markedly resistant to a wide range of conventional antibiotics, and are often related to multi-resistant strains. Therefore, an increasing number of scientific studies are documenting morphotypes of clinically significant bacteria.

In principle, colony morphology characterisation is a simple procedure and it is fairly easy to integrate into the analytical pipeline of any laboratory. Colony morphology is described through naked-eye observation and using a magnifying glass, and classified

* Corresponding author at: ESEI: Escuela Superior de Ingeniería Informática, University of Vigo, Edificio Politécnico, Campus Universitario As Lagoas s/n, 32004 Ourense, Spain.

E-mail addresses: anamargaridasousa@deb.uminho.pt (A.M. Sousa), mopereira@deb.uminho.pt (M.O. Pereira), analia@uvigo.es, analia@ceb.uminho.pt (A. Lourenço).

http://dx.doi.org/10.1016/j.jbi.2015.03.007
1532-0464/

using several criteria, commonly accepted in clinical microbiology. However, experimental design and morphological annotation should be consistent in order to allow the systematic comparison of morphotypes across experiments (and laboratories), species, diseases and clinical samples. Colony morphologies vary widely, depending on the particular behaviour of the microbial species under different test conditions (e.g. different colonisation sites, or antibiotic agents with different modes of action). Also, the morphological traits exhibited by the colonies may be significantly affected by the procedures taken to isolate and grow the bacteria [21]. Thus, it is very important to establish the minimum set of information to be part of the morphotype description and to employ harmonised vocabulary in both the biological contextualisation and the morphological characterisation of the observed colony.

Scientific literature is the main source of morphotypes, where they are often presented as exemplificative figures of what the researchers observed and are described informally. The fact that the description of colony morphologies does not yet follow predefined rules of annotation nor makes use of controlled vocabulary hampers the automated classification, integration and interpretation of such data.

Researchers are in need of new resources and tools geared to systematically analyse morphotypes, across infections, body locations, antimicrobial treatments and a number of other conditions of clinical relevance. Multiple ontologies have been proposed in the domain of phenotypes. Some are specialised in the characterisation of species (typically, model organisms), such as the Mammalian Phenotype ontology (MP) [22], the Worm Phenotype ontology (WPO) [23], the Plant ontology (PO) [24] and the Human Phenotype ontology (HPO) [25]. Others, like the Phenotype and Trait ontology (PATO) [26], are focused on integrating phenotypes across species, and reuse anatomy and process ontologies.

This paper presents the rationale of a novel ontological framework in support of the characterisation of the colony morphology of clinically significant bacteria. The Colony Morphology Ontology (CMO) is introduced as an integrative resource for the systematic, transparent and unambiguous characterisation of colony morphology traits in support of clinical diagnosis. CMO supports the MorphoCol database, a public Web repository of pathogenic bacterial morphotypes (http://morphocol.org). The aim of this repository is to enable the macroscopic observation of morphotypes and the comparison of the morphological "output" of the species in different scenarios (e.g. antibiotic therapeutics and body localisation).

The originality of this work lays on addressing colony morphotyping in a systematic, harmonised and computerised way. Although still in its infancy, MorphoCol aims to pave the way to the development of advanced clinical decision making applications, which may use morphological features as immediate indicators of microbial behaviour (Fig. 1). These indicators can be used to guide more sophisticated (time-consuming and costly) analyses, such as proteome and transcriptome analyses. Also, they may be used to construct decision support models that help clinicians in determining or anticipating what may be expected in terms of a given microbial species resistance and resilience in a clinical incident. To the best of our knowledge this is the first public repository documenting bacterial colony morphology systematically. Currently, the system documents respiratory infection traits. In the future, it will cover for other major infections regarding the urinary tract, bloodstream, chronic wounds, osteomyelitis and biomaterial-associated infections. Thus, MorphoCol will be of aid to the wider community of researchers and clinicians working in clinical microbiology.

2. Design and implementation

2.1. Considerations in ontology design

The CMO aims to characterise the main features of the morphology of bacterial colonies. Since most of the data is widespread in scientific literature, a curation pipeline based on expert manual annotation was implemented (Fig. 2). The documents are compiled through PubMed keyword-based searches combining terms related to morphological descriptions and bacterial colonies. The

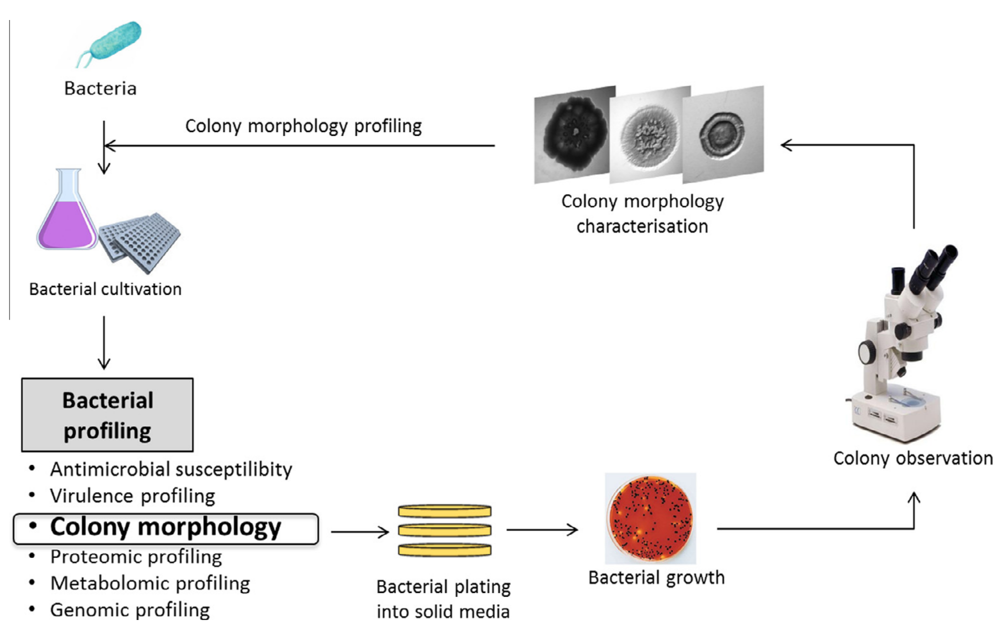

Fig. 1. General workflow of bacterial profiling. Bacteria provided from samples are cultivated and further characterised using several methods in which colony morphology is included. To perform colony morphology characterisation, bacteria from cultures is plated and grown onto solid media. Further, colonies are observed and characterised using 10 main morphological features: form, margin, sheath, type of surface, texture, consistency, opacity, size and colour. Likewise, colony morphotypes can be isolated and subjected to bacterial profiling using other methods, such as antimicrobial susceptibility and virulence characterisation.

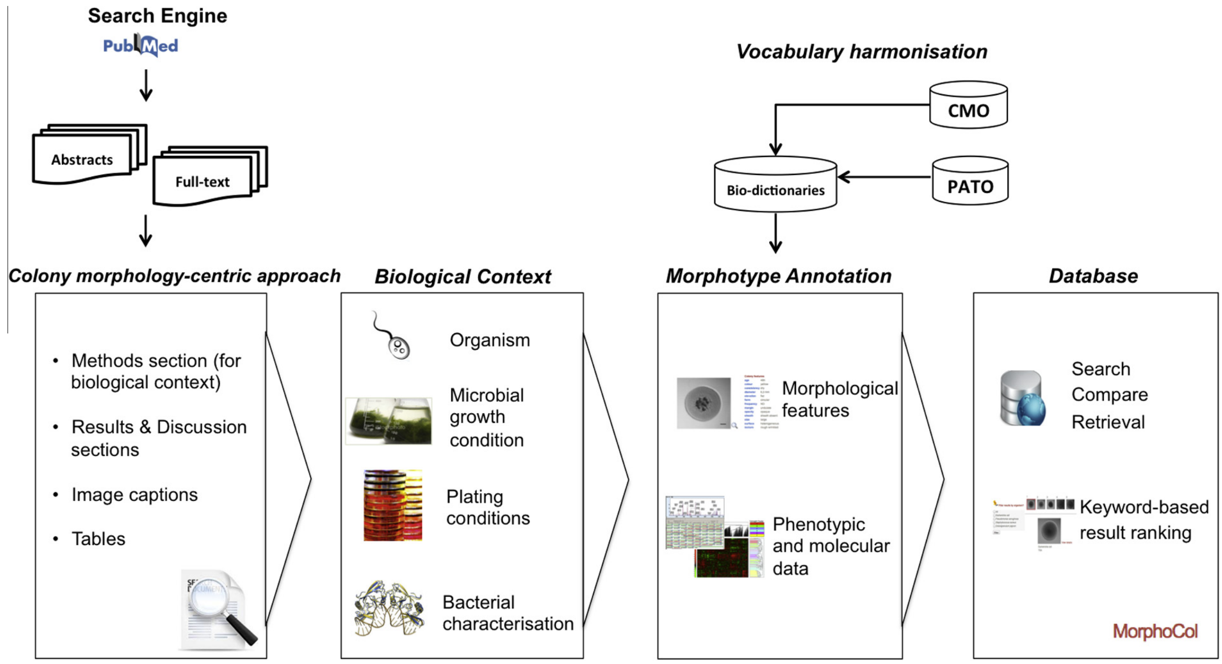

Fig. 2. An overview of the literature curation pipeline of MorphoCol. Colony morphotype curation starts with document retrieval via PubMed. Once all relevant information has been flagged, curators annotate the morphotypes using morphological features of colonies using CMO and PATO controlled vocabularies, and link morphotypes to the corresponding phenotypic and molecular data. MorphoCol search engine enables users to retrieve and compare morphologies, ranking results according to keywords of interest and grouped by bacterial species.

process of document retrieval encompasses the screening of abstracts and the download of the potentially relevant full-texts. Most of the contents are retrieved from the sections Materials & Methods and Results & Discussion (including the captions of existing figures and tables) of the reviewed documents.

It is equally important to identify the morphological descriptions as it is to characterise the biological context from where the morphotypes emerged. Most of the work of curators is centred on the preparation of an harmonised vocabulary that may support the systematic and comprehensive description of the morphotypes, namely: (1) collection of the terms commonly used by authors of clinical, microbiological and medical studies in the characterisation of colony morphology; (2) analysis of these textual descriptors, evaluating the appropriateness of the associated semantics and identifying the common name and synonyms of each concept according to overall concordance and our expertise in the field; and (3) manual validation of the descriptive ability of the set of concepts gathered against published descriptions of morphotypes.

Literature curation accounted for more than one hundred different terms, which were filtered out considering the exclusion criteria below:

– terms with no clear definition (for example, "normal colony"; "atypical morphology", "irregular shaped", and "normal size");
– terms referring to characteristics of bacteria-forming colony, i.e. characteristics of the bacteria that form the colony rather than the morphological features of the colony (for example, "rod-shaped bacteria");
– derived terms (for example, "semi-fluffy", "semi-dry", "degree of colour", "non-mucoid", "slightly rhizoid", "marginally convoluted");
– infrequent terms, i.e. those apparently used by only one author or research group.

Then, the remaining terms were checked for definition inconsistencies and term synonyms. Typically, the most used term was chosen as the main descriptor of the morphological feature and the other related terms were associated as synonyms. As a result, the structure of the CMO encompassed a total of 7 main categories and 33 sub-categories.

Harmonised and manual annotation guarantees the high quality of the morphotypes in MorphoCol repository. Likewise, it enables the search and comparison of morphotype according to various aspects of morphological and biological characterisations.

2.2. CMO format

The CMO was developed following the basic principles of the Open Biomedical Ontologies (OBO) Foundry [27]. In particular, the organisation of the CMO was based on the following main criteria:

– CMO is restricted to the morphological characterisation of bacterial colonies and, therefore, it contains just model concepts and relations that are relevant to the representation of colony data;
– CMO should be used for annotating data in databases and for textual documentation and as such, it should be understandable to people and unambiguously interpreted by software;
– CMO development follows a pragmatic approach that grants the ability to integrate new morphological descriptors as they arise without affecting the existing ontological structure;
– any bacterial colony morphology should be comprehensively described by a combination of CMO instances;
– whenever possible, CMO terms are cross-referenced to entries on other ontologies covering for phenotypic characterisation.

The ontology was constructed using the OBO-Edit editor, an open source platform that allows the editing of OBO-alike ontologies (http://oboedit.org/) [28]. Five pre-defined OBO tags were used to represent the CMO terms, including id, name, synonym, def and xref.

2.3. MorphoCol knowledgebase

MorphoCol is a publicly Web-accessible knowledgebase that documents bacterial colony morphotypes, as comprehensively as possible, in order to enable the search and comparison of morphotypes across species (and strains) and conditions (namely, diseases and colonisation sites). Currently, the main source of information is scientific literature and manual curation grants the high quality of the data available. The curation pipeline (Fig. 2) will be gradually incorporating automatic procedures, namely text mining processes, now that we have in hand an appropriate terminological resource and the minimum set of information necessary to comprehensively describe a morphotype. Likewise, the knowledgebase will enable the direct submission of morphotype data by authors, promoting a close interaction with the community.

MorphoCol server runs on a CentOS platform (version 5.6) with Apache HTTP server (version 2.2.22), MySQL Community Server (version 5.1.58) and PHP 5.5.3. Apache, MySQL and PHP technology are open-source and platform-independent software. Moreover, MySQL supports multi-threading and multi-user environments, and thus it is well-suited to support (increasing) real-world database usage. Currently, the Web server and all parts of the database are hosted at the Centre of Biological Engineering of the University of Minho, Portugal.

3. Results and discussion

Colony morphotyping is a common technique used in microbiological studies of varied purposes. It should be emphasised that this study does not propose any colony features. The colony morphology features mentioned and described are commonly accepted by the microbiological community and they can be traced back to as early as papers published decades ago [29–31].

3.1. CMO structure and contents

The validation of the CMO terms against published morphological descriptions showed that all terms included in the ontology structure were valid, but pointed out some additional considerations as follows: *Pseudomonas aeruginosa* colonies exhibit an enveloping part or structure after the margin that surrounds the colony, which is referred to as 'sheath' (Fig. 3A); colonies can exhibit more than one type of surface, e.g. *P. aeruginosa* colonies can exhibit surfaces with smooth and rough zones (Fig. 3B), smooth and wrinkled zones (Fig. 3C), or wrinkled and rough zones (Fig. 3D). To enhance its descriptive abilities, the initial structure of the CMO was extended to a total of 10 main categories and 37 sub-categories (Fig. 4).

An important issue whilst defining the organisation of the CMO was the ability to perform updates without causing major changes in the structure. To this end, the high level nodes of the ontology represent the general concepts behind the morphological features more frequently discussed in the literature, including 'form', 'margin', 'type of surface', 'texture', 'sheath', 'opacity', 'elevation', 'consistency', 'size' and 'colour'. The following subsections introduce these concepts, explaining the semantics adopted by the CMO and discussing the semantic inconsistencies and ambiguities found in the literature.

3.1.1. Form

The term 'form' is commonly used to describe the whole configuration of a colony. For example, "… overall surface shape (convex, crater, lobulated, radial, radioumbilicated, radio-umbonated, peaked, rugose, segmented-rugose, segmented-umbilicated, umbilicated, umbilicated with irregular edge, umbilicated with heaped-up irregular edge, or umbonated)." [32], or to differentiate colonies with certain characteristics, such as "… two distinct colony morphology variants that switched between a transparent form, facilitating adherence and carriage, and an opaque form that poorly adhered…" [33]. So, in the CMO, the term 'colony form' (CMO:0000001) is defined as the geometrical configuration of the colony – "a morphological quality inhering in a colony by virtue of having a configuration".

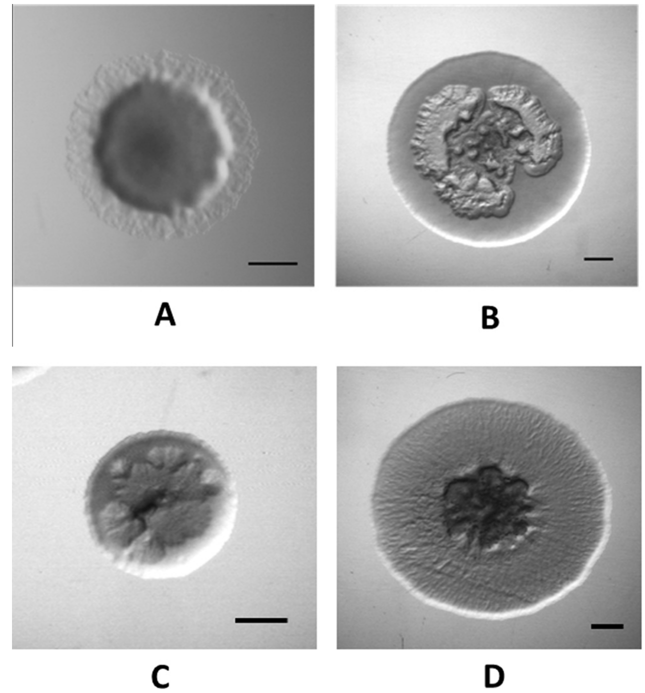

Fig. 3. Examples of colony morphologies exhibiting features that the first version of the CMO did not consider: (A) sheath, enveloping part or structure after the margin that surrounds the colony; dual type of textures, (B) smooth and wrinkled, (C) smooth and rough, and (D) rough and wrinkled, described from the periphery to the centre. Black bar = 1 mm.

3.1.2. Margin

The description of the margin of the colony, also referred to as edge or border, is typically based on the characteristics of the colony circumference. For example, "Colony type 2 was hard, more orange in colour, non-rhizoid or only slightly rhizoid, and had irregular edges and convex growth form. Colony type 3 had round edges, and smooth, yellowish appearance." [34], "'fried egg' SCVs, with translucent edges surrounding a smaller elevated…" [19], "other colonies (approximately 60% of the total) were smaller with somewhat rough edges" [35]. In the CMO, the term 'colony margin' (CMO:0000002) represents the configuration of the limiting border of the colony and is described as "a morphological quality inhering in a colony by virtue of having a limit zone".

3.1.3. Sheath

The description of 'colony sheath' is not common in colony morphology observation. However, this structure is often present in *P. aeruginosa* colonies, which are clinically recurrent, and it is quite variable (Fig. 3). Based on our expertise, this may be a valuable element in the description of colony differentiation. So, it was included in the CMO (CMO:0000005) and defined as "a morphological quality inhering in a colony by virtue of having a closely enveloping part or structure after the margin and around the colony".

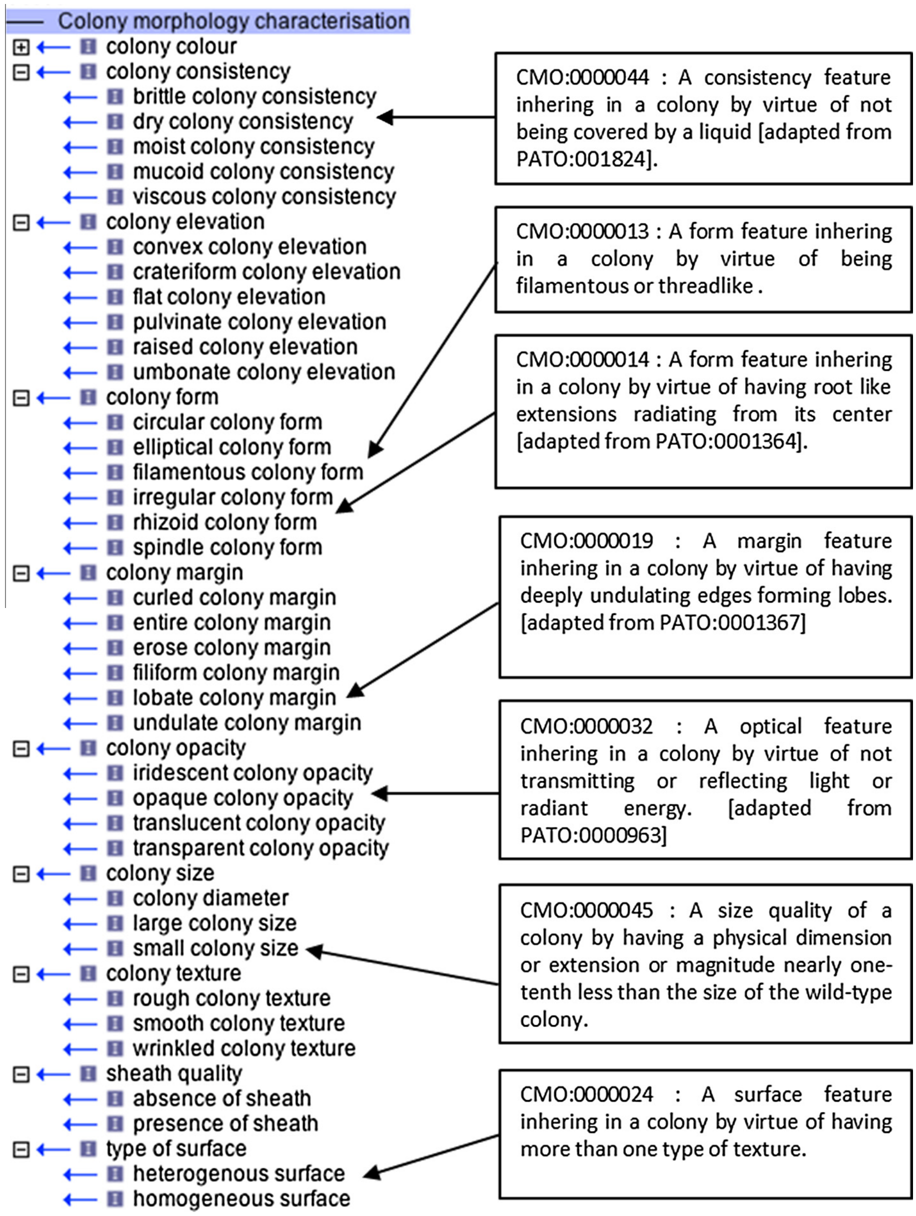

Fig. 4. Structure of the CMO with definitions of some terms directly imported or adapted from the PATO.

3.1.4. Type of surface

As abovementioned, some bacterial species can form colonies with more than one type of texture at the surface (Fig. 3B–D). As such, a new category, named 'type of surface' (CMO:0000003), was added to the ontology to describe the type of textures or irregularities exhibited by the colony surface. Notably, 'homogeneous surface' (CMO:0000023) is a surface with just one type of texture, and 'heterogeneous surface' (CMO:0000024), is a surface that presents more than one type of texture.

3.1.5. Texture

Colony texture, also referred to as surface or roughness, is a feature very common in morphological descriptions. Its importance arises from experimental evidence that rough colonies are often associated with key pathogenic phenomena, such as augmented virulence potential [36,37].

In the literature, we may find examples of general texture descriptions applied to colony morphology such as: ''PA14 lasR mutant formed a flat, smooth colony as compared to the wrinkled wild-type phenotype'' [36], ''the ST variant formed colonies that were smaller and had a rough, wrinkled, and dry surface appearance compared to the smooth, larger WT colonies'' [38] and ''... type I (wrinkled, purple, dry and irregular giant cristae; $n = 8$), II (wrinkled, purple, dry and volcanolike; $n = 6$), (...) V (smooth, pale, mucoid and raised circles; $n = 3$), VI (smooth, pale, mucoid and even glistening; $n = 2$)...'' [39]. This feature may also be used to describe the form of the colonies. For example, ''... overall surface shape (convex, crater, lobulated, radial, radio-umbilicated, radio-umbonated, peaked, rugose, segmented-rugose, segmented-umbilicated, umbilicated, umbilicated with irregular edge, umbilicated with heaped-up irregular edge, or umbonated)'' [32]. And, often enough, the types of texture rough and wrinkled are used as synonymous or as discriminating characteristics [37,38,40].

The CMO addresses this conceptual ambiguity by defining the terms 'wrinkled' and 'rough' in accordance to the interpretation provided by the majority of authors, which is consistent with our expertise. The term 'colony texture' (CMO:0000004) denotes the presence or absence of irregularities on colony surface. When irregularities are present, the form is depicted as 'wrinkled colony texture' (CMO:0000027), and if the vertical irregularities are large or small it is classified as 'rough colony texture' (CMO:0000026). A surface showing no irregularities is described as 'smooth colony texture' (CMO:0000025).

3.1.6. Elevation

Typically, the term 'elevation' describes the form of growth of the colony observed from a side perspective. For example, "Colony type 1 was rhizoid and flat with yellow centre. Colony type 2 was hard, more orange in color, non-rhizoid or only slightly rhizoid, and had irregular edges and convex growth form." [34]. In the CMO, the terms 'colony elevation' (CMO:0000007) and 'colony form' (CMO:0000001) are perfectly differentiated, that is one corresponds to the side perspective of colony form and the other describes the overall geometric colony configuration, respectively.

3.1.7. Consistency

The term 'consistency' is frequently referred as the wetness or mucoidy of the colony [33,37]. The characterisation of colony consistency gained renewed importance due to the impact of mucoid colonies in cystic fibrosis lung disease [41–43]. For example, "... characterisation of two *Burkholderia cenocepacia* sequential isolates displaying different morphotypes (mucoid vs. non-mucoid) isolated from a CF patient..." [44], "... *P. aeruginosa* PDO300 showed a nonmucoid colony morphology..." [45] and "... type I (wrinkled, purple, dry and irregular giant cristae; $n = 8$) (...) IV (wrinkled, pale, semi-dry and volcanolike; $n = 2$), V (smooth, pale, mucoid and raised circles; $n = 3$)..." [39]. Similarly to what happens with the term 'texture' (CMO:0000004), the term 'colony consistency' (CMO:0000008) is often described as a form characteristic [32]. In the CMO, the terms 'colony form' and 'colony consistency' were separated and 'colony consistency' was defined as "a physical quality inhering in a colony by virtue of having density, firmness, or viscosity".

3.1.8. Opacity

Opacity has been described as an important colony feature in different pathogenic morphotypes. For instance, in the case of *Streptococcus pneumonia* colonies "... two distinct colony morphology variants that switched between a transparent form, facilitating adherence and carriage, and an opaque form that poorly adhered..." [33] and in the case of *Pseudomonas fluorescens* colonies, "... The original isolate formed thick opaque colonies...", "... After three days of growth in liquid KB medium, flat and translucent colonies were found..." [9]. Therefore, the term 'colony opacity' (CMO:0000006) was included in the CMO and defined as "an optical quality which obtains by virtue of the ability of the mass of the colony to absorb visible light".

3.1.9. Size

The feature 'size' is a well-recognised trait in colony characterisation. For instance: "... strain 43895OR is also similar to the rdar strains of *Salmonella enterica* serovar Typhimurium in characteristics such as colony size...", "... *Escherichia coli* strain 43895OR forms a larger colony..." [46], "... Temperature-induced SCVs were <1 mm in colony size..." [47] and "... Analysing the persisting bacteria revealed a high phenotypic diversity, showing normal, small and very small colonies..." [48].

The significance of the size description arises from the association between the appearance of small colonies and increased antibiotic resistance [19,49]. However, it has conceptual gaps. For instance, Haussler and colleagues [50] considered that *P. aeruginosa* small colonies are those that have 1–3 mm of diameter. In contrast, several other authors study *P. aeruginosa* small colonies without any assumption about the diameter range of the colony. Additionally, this definition of size is highly taxon-dependent. For instance, the dimension of a small colony of *P. aeruginosa*, in general defined as below 3 mm, does not correspond to the dimension of a small colony of *Staphylococcus aureus*, typically defined as below 1 mm. In the CMO, the child terms of 'colony size' (CMO:0000009) were defined in a taxon-independent manner such that the term 'small colony size' (CMO:0000045) denotes "a size quality of a colony by having a physical dimension or extension or magnitude nearly one-tenth less than the size of the wild-type colony" and the term 'large colony size' (CMO:0000046) denotes colonies exceeding this threshold.

Diameter measurements are also frequently used to characterise the size of the colony. For instance, "... The following features of colony morphology were recorded: colony size (measured in mm by ruler)..." [33] and "... cells of the rdar strains were reported to reach colony diameters of 6 cm..." [46]. In the CMO, the term 'colony diameter' (CMO:0000047) was included as a child term of 'colony size' (CMO:0000009) and defined as the distance between two equidistant points of the margin or sheath. Typically, colony diameter is calculated manually using a rule. The use of image software tools, such as ImageJ [51], could be of assistance (systematic and fast large-scale calculation), but issues relating to region segmentation and size computing methods need to be discussed first.

3.1.10. Colour

The chromogenesis or colour exhibited by bacterial colonies is very important when describing clinical morphotypes because it is associated with the differential production of various pigments [52,53]. Some examples of colour description applied to colony morphology are: "... certain strains of *Salmonella* that displayed a red and dry colony phenotype..."; "... the white variants do not revert back to a dry, red phenotype..." [46], "... StNMSm (straw-colored, nonmucoid, and smooth) morphotypes exhibited decreased frequency of pyocyanin overproduction..." [52] and "Colony type 3 had round edges, and smooth, yellowish appearance. Type 4 colonies were white or light yellow, smooth and spreading on the agar with irregular shape." [34]. In the CMO, the term 'colony colour' (CMO:0000010) was defined as "a composite chromatic quality composed of hue, saturation and intensity parts". A diverse and automatic *pallete* of colours could be incorporated under the term 'colour', but we chose to focus on the colours that have already been documented in the literature. The CMO includes these "basic" colours, notably 'white' (CMO:0000048), 'black' (CMO:0000049), 'grey' (CMO:0000050), 'brown' (CMO:0000051), 'yellow' (CMO:0000052), 'red' (CMO:0000053), 'orange' (CMO:0000054), 'green' (CMO:0000055), 'blue' (CMO:0000056), 'pink' (CMO:0000057) and 'violet' (CMO:0000058).

3.2. PATO as ontological reference

The analysis of existing phenotypic ontologies, such as PATO [26], MP [22], WPO [23], PO [24] and HPO [25], revealed a number of terms in common. However, PATO is the only ontology that provides taxon-independent and general phenotypic descriptors, which is one of the main requirements of the CMO design.

No CMO term is defined identically to a PATO term. A total of 43 terms of PATO (Fig. 4) were adapted to fit the CMO domain. For example, the term 'circular' in PATO is defined as "a shape quality inhering in a bearer by virtue of the bearer's being such that every part of the surface or the circumference is equidistant from the center" and in CMO is defined as "a form feature inhering in a colony by virtue to present a configuration of a circumference or a circle due to any point of the edge be equidistant from the center". So, the labels of the CMO terms were altered to prevent any confusion about the CMO and PATO terms and their definitions. For example, 'circular colony form' (CMO:0000011), 'smooth colony texture' (CMO:0000025), 'transparent colony opacity' (CMO:0000030), 'flat colony elevation' (CMO:0000034), 'viscous colony consistency' (CMO:0000041), and 'colony colour' (CMO:0000010). The adaptation of terms and definitions was also motivated by differences in the organisation of the PATO and the CMO, and affected 14 terms. For instance, the term 'lobate' (PATO:0001367) in the PATO is "a surface feature shape quality inhering in a bearer by virtue of the bearer's having deeply undulating edges forming lobes" whilst the term 'lobate colony margin' (CMO:0000019) in the CMO is not a surface feature shape quality, but rather a margin feature quality.

A total of 10 CMO terms did not find any correspondence with PATO due to semantics differences. For example, the term 'small' (PATO:0000587) is described in the PATO as "a size quality which is relatively low", but in the scope of bacterial colonies the definition of 'small colony size' (CMO:0000045) is more specific, i.e. "a physical magnitude 10 times smaller than the diameter of the wild-morphotype".

3.3. Availability

The CMO ontology files, including terms, definitions and relationships, are freely available at the MorphoCol Web knowledgebase for bacterial colony morphotypes (http://morphocol.org) and the portal of the international consortium Minimum Information About a Biofilm Experiment (MIABiE) (http://miabie.org).

3.4. Web application – MorphoCol knowledgebase

The volume of phenotypic data on bacterial colonies is growing considerably due to the use of high-throughput experimental procedures and analytical methods, and the impressive ability of bacteria to diversify under several environmental conditions. The variety of morphotypes within and across species is great, and requires expert knowledge on the formation of colonies and on the pathogenic bacteria under study in order to describe the observed morphological features comprehensively. Computer applications in assistance of phenotypic annotation and analysis are therefore considered pivotal to manage and integrate data in an amenable, systematic and accurate way. In particular, biomedical ontologies are the key to standardise terminology and describe relevant morphological features unambiguously.

The immediate application of the CMO is the sharing of current understanding of the variation of colony morphology in microbial infection among domain experts, both clinicians and researchers. Furthermore, the CMO may be of help to information retrieval applications, providing vocabulary and taxonomy that can be used for query expansion and semantic searching in this domain.

The MorphoCol knowledgebase is supported by the CMO and aims to help manage the fast proliferation of information about colony morphology. Currently, documentation efforts are focused on morphotypes exhibited by pathogenic bacteria causing respiratory infections, one of the most prevalent types of infection worldwide. Typically, respiratory human pathogens, such as *P. aeruginosa* and *S. aureus*, differentiate when colonising the human lungs [54]. For instance, several colony morphologies, such as the *P. aeruginosa* mucoid morphotype [42] and small colony variants of *S. aureus* [55], have been correlated to cystic fibrosis development. The identification of these and other clinically significant colony morphotypes is of tremendous importance because colony observation is quite immediate and costless when compared to state-of-the-art identification methods. Even without pinpointing the strains involved, morphotype comparison may provide insights on the stage of infection (acute, intermediate or chronic) and the resistance and virulence levels to be expected from the bacteria.

As such, colony morphologies are to be documented with various descriptive metadata (Fig. 5), including information about the experimental conditions and the morphology observed. Experimental information is related to the "circumstances" in which colonies are collected, e.g. the species and strains involved, the type of culture that bacteria are coming from (e.g. planktonic or biofilm), and the plating conditions in which bacteria were allowed to grow in solid media to form colonies (e.g. medium, time of growth, temperature, respiratory conditions). The description of such information is of upmost importance since experimental conditions significantly affect the colony morphogenesis [21]. The morphological annotation is supported by the CMO. The experiment is profiled in terms of authorship (institution and person of contact), the data scope, the date of execution, and derived publications. The morphotype data record is manually verified in order to ensure its quality and thus allow clinicians and researchers to consider such data in their analyses and the formulation of new hypotheses.

4. Conclusions and challenges

Several authors have documented colony morphologies in clinical settings and have shown that morphological features may be indicative of underlying microbial cues, and most notably, of resistance and virulence responses [55,56]. Over the years, this research team has analysed a large number of colony morphologies generated by clinically significant bacteria and developed expertise on colony observation and annotation. This led to the development of a specialised method of analysis that aims to deliver useful inputs to more elaborated (and costly) studies, and assistance to clinical decision making [21]. The development of the CMO, the first ever controlled vocabulary on colony morphology, and the MorphoCol knowledgebase are considered an important step forward for enabling the standardised and systematic annotation of morphotypes. For the first time, there is a knowledgebase dedicated to the management of data related to colony morphotypes, including morphological data and experimental metadata. This knowledgebase provides the basic means to enquire and compare the visual manifestations of bacterial evolution and adaptation processes across pathogenic microorganisms and infections.

In the short term, efforts will be focused on extending the description of the pathogenic potential of colony-forming bacteria, particularly regarding the expression of virulence factors, such as the ability to form biofilms, the production of toxins and quorum-sensing molecules. Moreover, the query tool will be complemented by a customisable comparison tool that looks for morphological similarities across species, infection sites and diseases. Therefore, the tool will provide insights on the most relevant traits of colony morphology under a given clinical scenario, which may be useful as predictive features of the virulence potential and resistance profile of bacteria causing the infection.

MorphoCol welcomes contributions from any research group working on the characterisation of clinically relevant morphotypes,

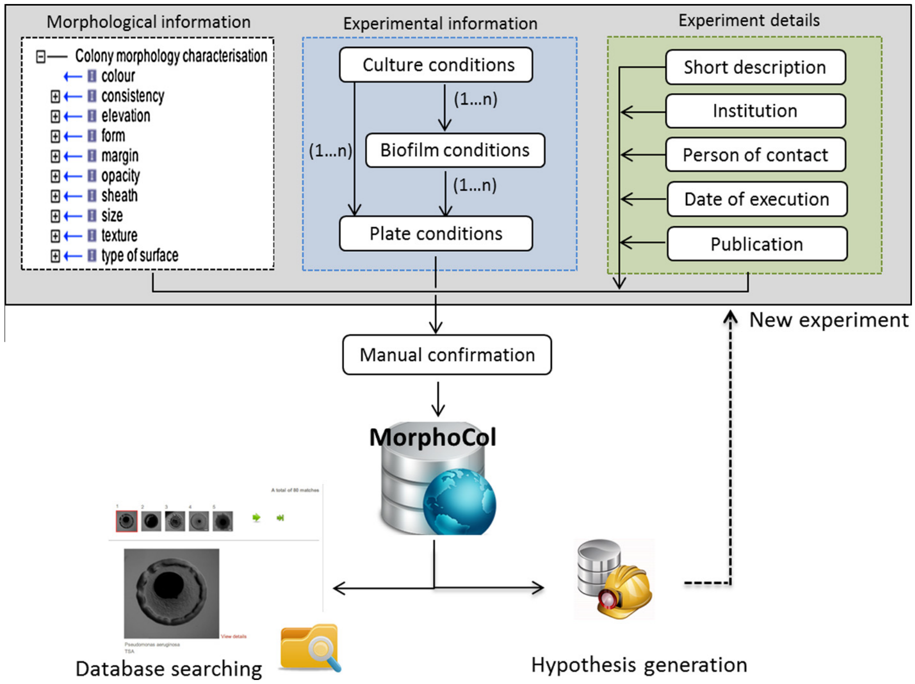

Fig. 5. The flowchart of the MorphoCol annotation process. Colony data annotation consists of the curation of morphological, experimental metadata and details followed by manual confirmation. After validation, morphotype data record is available online for information retrieval and formulation of new hypotheses.

crediting data authorship and associated bibliographic references. Currently, the system documents respiratory infection traits. In the near future, it will cover for other major infections regarding the urinary tract, bloodstream, chronic wounds, osteomyelitis and biomaterial-associated infections. Thus, MorphoCol will be of aid to the wider community of researchers and clinicians working in clinical microbiology.

Given that the ultimate aim is to rationalise morphological and biological associations within and across species, and across infections, molecular, proteomic, transcriptomic and genomic profiles will be a key, future asset. For instance, proteomic approaches may support the identification of morphotypes-biomarkers and future target sites for new drugs whilst transcriptomic approaches may provide snapshots of gene expression during infection development and microbial response to antimicrobial treatments. These data will validate the correlation of morphotypes to resistance and virulence cues, and enhance our understanding about the different responses of pathogens to antimicrobial therapies in acute and chronic infections. Altogether, these data will provide valuable inputs to clinical research and decision-making.

Competing interests

The authors declare that they have no competing interests.

Acknowledgments

The authors thank the project PTDC/SAU-ESA/646091/2006/FCOMP-01-0124-FEDER-007480FCT, the Strategic Project PEst-OE/EQB/LA0023/2013, the Project “BioHealth – Biotechnology and Bioengineering approaches to improve health quality”, Ref. NORTE-07-0124-FEDER-000027, co-funded by the Programa Operacional Regional do Norte (ON.2 – O Novo Norte), QREN, FEDER, the project “RECI/BBB-EBI/0179/2012 – Consolidating Research Expertise and Resources on Cellular and Molecular Biotechnology at CEB/IBB”, Ref. FCOMP-01-0124-FEDER-027462, FEDER, and the Agrupamento INBIOMED from DXPCTSUG-FEDER unha maneira de facer Europa (2012/273). The research leading to these results has received funding from the European Union’s Seventh Framework Programme FP7/REGPOT-2012-2013.1 under grant agreement n° 316265, BIOCAPS. This document reflects only the author’s views and the European Union is not liable for any use that may be made of the information contained herein. The authors also acknowledge PhD Grant of Ana Margarida Sousa SFRH/BD/72551/2010.

References

[1] J. Weile, C. Knabbe, Current applications and future trends of molecular diagnostics in clinical bacteriology, Anal. Bioanal. Chem. 394 (2009) 731–742.
[2] A. van Belkum, G. Durand, M. Peyret, S. Chatellier, G. Zambardi, J. Schrenzel, D. Shortridge, A. Engelhardt, W.M. Dunne, Rapid clinical bacteriology and its future impact, Ann. Lab. Med. 33 (2013) 14–27.
[3] M. Welker, Proteomics for routine identification of microorganisms, Proteomics 11 (2011) 3143–3153.
[4] P.K. Mandal, A.K. Biswas, K. Choi, U.K. Pal, Methods for rapid detection of foodborne pathogens – an overview, Am. J. Food Technol. 6 (2011) 87–102.
[5] S.Y. Hsieh, C.L. Tseng, Y.S. Lee, A.J. Kuo, C.F. Sun, Y.H. Lin, J.K. Chen, Highly efficient classification and identification of human pathogenic bacteria by MALDI-TOF MS, Mol. Cell. Proteomics 7 (2008) 448–456.
[6] S. Yachi, M. Loreau, Biodiversity and ecosystem productivity in a fluctuating environment: the insurance hypothesis, Proc. Natl. Acad. Sci. USA 96 (1999) 1463–1468.
[7] B.R. Boles, M. Thoendel, P.K. Singh, Self-generated diversity produces ‘insurance effects’ in biofilm communities, Proc. Natl. Acad. Sci. USA 101 (2004) 16630–16635.
[8] C. Goerke, M. Gressinger, K. Endler, C. Breitkopf, K. Wardecki, M. Stern, C. Wolz, B.C. Kahl, High phenotypic diversity in infecting but not in colonizing *Staphylococcus aureus* populations, Environ. Microbiol. 9 (2007) 3134–3142.
[9] G. Rossignol, D. Sperandio, J. Guerillon, C. Duclairoir Poc, E. Soum-Soutera, N. Orange, M.G. Feuilloley, A. Merieau, Res. Microbiol. 160 (2009) 337–344.
[10] J.A. Davies, J.J. Harrison, L.L. Marques, G.R. Foglia, C.A. Stremick, D.G. Storey, R.J. Turner, M.E. Olson, H. Ceri, The GacS sensor kinase controls phenotypic reversion of small colony variants isolated from biofilms of *Pseudomonas aeruginosa* PA14, FEMS Microbiol. Ecol. 59 (2007) 32–46.
[11] T. Tannaes, H.J. Grav, G. Bukholm, Lipid profiles of *Helicobacter pylori* colony variants, APMIS: Acta Pathol. Microbiol. Immunol. Scand. 108 (2000) 349–356.
[12] D.W. Martin, M.J. Schurr, M.H. Mudd, J.R. Govan, B.W. Holloway, V. Deretic, Mechanism of conversion to mucoidy in *Pseudomonas aeruginosa* infecting cystic fibrosis patients, Proc. Natl. Acad. Sci. USA 90 (1993) 8377–8381.
[13] R.C. Massey, A. Buckling, S.J. Peacock, Phenotypic switching of antibiotic resistance circumvents permanent costs in *Staphylococcus aureus*, Curr. Biol. 11 (2001) 1810–1814.

[14] K. Lewis, Persister cells and the riddle of biofilm survival, Biochemistry-Moscow 70 (2005) 267–274.
[15] A.M. Sousa, I. Machado, M.O. Pereira, Phenotypic switching: an opportunity to bacteria thrive, in: A. Mendez-Vilas (Ed.), Science against Microbial Pathogens: Communicating Current Research and Technological Advances, Formatex Research Center, Spain, 2011.
[16] G. Agarwal, A. Kapil, S.K. Kabra, B.K. Das, S.N. Dwivedi, Characterization of *Pseudomonas aeruginosa* isolated from chronically infected children with cystic fibrosis in India, BMC Microbiol. 5 (2005) 43.
[17] G. Manno, M. Cruciani, L. Romano, S. Scapolan, M. Mentasti, R. Lorini, L. Minicucci, Antimicrobial use and *Pseudomonas aeruginosa* susceptibility profile in a cystic fibrosis centre, Int. J. Antimicrob. Agents 25 (2005) 193–197.
[18] N. Wellinghausen, I. Chatterjee, A. Berger, A. Niederfuehr, R.A. Proctor, B.C. Kahl, Characterization of clinical *Enterococcus faecalis* small-colony variants, J. Clin. Microbiol. 47 (2009) 2802–2811.
[19] R.A. Proctor, C. von Eiff, B.C. Kahl, K. Becker, P. McNamara, M. Herrmann, G. Peters, Small colony variants: a pathogenic form of bacteria that facilitates persistent and recurrent infections, Nat. Rev. Microbiol. 4 (2006) 295–305.
[20] Q. Wei, S. Tarighi, A. Dotsch, S. Haussler, M. Musken, V.J. Wright, M. Camara, P. Williams, S. Haenen, B. Boerjan, et al., Phenotypic and genome-wide analysis of an antibiotic-resistant small colony variant (SCV) of *Pseudomonas aeruginosa*, PLoS ONE 6 (2011) e29276.
[21] A.M. Sousa, I. Machado, A. Nicolau, M.O. Pereira, Improvements on colony morphology identification towards bacterial profiling, J. Microbiol. Methods 95 (2013) 327–335.
[22] C.L. Smith, C.A. Goldsmith, J.T. Eppig, The mammalian phenotype ontology as a tool for annotating, analyzing and comparing phenotypic information, Genome Biol. 6 (2005) R7.
[23] G. Schindelman, J.S. Fernandes, C.A. Bastiani, K. Yook, P.W. Sternberg, Worm phenotype ontology: integrating phenotype data within and beyond the *C. elegans* community, BMC Bioinformatics 12 (2011) 32.
[24] Y. Yamazaki, P. Jaiswal, Biological ontologies in rice databases. An introduction to the activities in Gramene and Oryzabase, Plant Cell Physiol. 46 (2005) 63–68.
[25] P.N. Robinson, S. Kohler, S. Bauer, D. Seelow, D. Horn, S. Mundlos, The human phenotype ontology: a tool for annotating and analyzing human hereditary disease, Am. J. Hum. Genet. 83 (2008) 610–615.
[26] G.V. Gkoutos, E.C. Green, A.M. Mallon, A. Blake, S. Greenaway, J.M. Hancock, D. Davidson, Ontologies for the description of mouse phenotypes, Comp. Funct. Genomics 5 (2004) 545–551.
[27] B. Smith, M. Ashburner, C. Rosse, J. Bard, W. Bug, W. Ceusters, L.J. Goldberg, K. Eilbeck, A. Ireland, C.J. Mungall, et al., The OBO foundry: coordinated evolution of ontologies to support biomedical data integration, Nat. Biotechnol. 25 (2007) 1251–1255.
[28] J. Day-Richter, M.A. Harris, M. Haendel, S. Lewis, OBO-Edit – an ontology editor for biologists, Bioinformatics 23 (2007) 2198–2200.
[29] E. Ben-Jacob, I. Cohen, D.L. Gutnick, Cooperative organization of bacterial colonies: from genotype to morphotype, Annu. Rev. Microbiol. 52 (1998) 779–806.
[30] C.H. Zierdt, P.J. Schmidt, Dissociation in *Pseudomonas aeruginosa*, J. Bacteriol. 87 (1964) 1003–1010.
[31] R.C. Clowes, D. Rowley, Genetic studies on small-colony variants of *Escherichia coli* K-12, J. Gen. Microbiol. 13 (1955) 461–473.
[32] N. Chantratita, V. Wuthiekanun, K. Boonbumrung, R. Tiyawisutsri, M. Vesaratchavest, D. Limmathurotsakul, W. Chierakul, S. Wongratanacheewin, S. Pukritiyakamee, N.J. White, et al., Biological relevance of colony morphology and phenotypic switching by *Burkholderia pseudomallei*, J. Bacteriol. 189 (2007) 807–817.
[33] M. Allegrucci, K. Sauer, Characterization of colony morphology variants isolated from *Streptococcus pneumoniae* biofilms, J. Bacteriol. 189 (2007) 2030–2038.
[34] H.M. Kunttu, L.R. Suomalainen, E.I. Jokinen, E.T. Valtonen, Flavobacterium columnare colony types: connection to adhesion and virulence?, Microb Pathog. 46 (2009) 21–27.
[35] D. Neut, J.G.E. Hendriks, J.R. van Horn, H.C. van der Mei, H.J. Busscher, *Pseudomonas aeruginosa* biofilm formation and slime excretion on antibiotic-loaded bone cement, Acta Orthop. 76 (2005) 109–114.
[36] R. Gupta, M. Schuster, Quorum sensing modulates colony morphology through alkyl quinolones in *Pseudomonas aeruginosa*, BMC Microbiol. 12 (2012) 30.
[37] M. Starkey, J.H. Hickman, L. Ma, N. Zhang, S. De Long, A. Hinz, S. Palacios, C. Manoil, M.J. Kirisits, T.D. Starner, et al., *Pseudomonas aeruginosa* rugose small-colony variants have adaptations that likely promote persistence in the cystic fibrosis lung, J. Bacteriol. 191 (2009) 3492–3503.
[38] M.J. Kirisits, L. Prost, M. Starkey, M.R. Parsek, Characterization of colony morphology variants isolated from *Pseudomonas aeruginosa* biofilms, Appl. Environ. Microbiol. 71 (2005) 4809–4821.
[39] Y.S. Chen, H.H. Lin, C.C. Hung, J.J. Mu, Y.S. Hsiao, Y.L. Chen, Phenotypic characteristics and pathogenic ability across distinct morphotypes of *Burkholderia pseudomallei* DT, Microbiol. Immunol. 53 (2009) 184–189.
[40] L. Friedman, R. Kolter, Genes involved in matrix formation in *Pseudomonas aeruginosa* PA14 biofilms, Mol. Microbiol. 51 (2004) 675–690.
[41] R.B. Troxler, W.C. Hoover, L.J. Britton, A.M. Gerwin, S.M. Rowe, Clearance of initial mucoid *Pseudomonas aeruginosa* in patients with cystic fibrosis, Pediatr. Pulmonol. 47 (2012) 1113–1122.
[42] O. Ciofu, L.F. Mandsberg, H. Wang, N. Hoiby, Phenotypes selected during chronic lung infection in cystic fibrosis patients: implications for the treatment of *Pseudomonas aeruginosa* biofilm infections, FEMS Immunol. Med. Microbiol. 65 (2012) 215–225.
[43] P. Greally, P. Whitaker, D. Peckham, Challenges with current inhaled treatments for chronic *Pseudomonas aeruginosa* infection in patients with cystic fibrosis, Curr. Med. Res. Opin. 28 (2012) 1059–1067.
[44] I.N. Silva, A.S. Ferreira, J.D. Becker, J.E. Zlosnik, D.P. Speert, J. He, D. Mil-Homens, L.M. Moreira, Mucoid morphotype variation of *Burkholderia multivorans* during chronic cystic fibrosis lung infection is correlated with changes in metabolism, motility, biofilm formation and virulence, Microbiology 157 (2011) 3124–3137.
[45] I.D. Hay, U. Remminghorst, B.H. Rehm, MucR, a novel membrane-associated regulator of alginate biosynthesis in *Pseudomonas aeruginosa*, Appl. Environ. Microbiol. 75 (2009) 1110–1120.
[46] G.A. Uhlich, P.H. Cooke, E.B. Solomon, Analyses of the red-dry-rough phenotype of an *Escherichia coli* O157:H7 strain and its role in biofilm formation and resistance to antibacterial agents, Appl. Environ. Microbiol. 72 (2006) 2564–2572.
[47] L.A. Onyango, R.H. Dunstan, J. Gottfries, C. von Eiff, T.K. Roberts, Effect of low temperature on growth and ultra-structure of *Staphylococcus* spp, PLoS ONE 7 (2012) e29031.
[48] L. Tuchscherr, E. Medina, M. Hussain, W. Volker, V. Heitmann, S. Niemann, D. Holzinger, J. Roth, R.A. Proctor, K. Becker, et al., *Staphylococcus aureus* phenotype switching: an effective bacterial strategy to escape host immune response and establish a chronic infection, EMBO Mol. Med. 3 (2011) 129–141.
[49] C. von Eiff, *Staphylococcus aureus* small colony variants: a challenge to microbiologists and clinicians, Int. J. Antimicrob. Agents 31 (2008) 507–510.
[50] S. Haussler, B. Tummler, H. Weissbrodt, M. Rohde, I. Steinmetz, Small-colony variants of *Pseudomonas aeruginosa* in cystic fibrosis, Clin. Infect. Dis.: Off. Publ. Infect. Dis. Soc. Am. 29 (1999) 621–625.
[51] S.M. Hartig, Basic image analysis and manipulation in ImageJ, in: Curr. Protoc. Mol. Biol., John Wiley & Sons, Inc., 2001.
[52] E. Mowat, S. Paterson, J.L. Fothergill, E.A. Wright, M.J. Ledson, M.J. Walshaw, M.A. Brockhurst, C. Winstanley, *Pseudomonas aeruginosa* population diversity and turnover in cystic fibrosis chronic infections, Am. J. Respir. Crit. Care Med. 183 (2011) 1674–1679.
[53] J.L. Fothergill, E. Mowat, M.J. Ledson, M.J. Walshaw, C. Winstanley, Fluctuations in phenotypes and genotypes within populations of *Pseudomonas aeruginosa* in the cystic fibrosis lung during pulmonary exacerbations, J. Med. Microbiol. 59 (2010) 472–481.
[54] A.R. Hauser, M. Jain, M. Bar-Meir, S.A. McColley, Clinical significance of microbial infection and adaptation in cystic fibrosis, Clin. Microbiol. Rev. 24 (2011) 29–70.
[55] S. Yagci, G. Hascelik, D. Dogru, U. Ozcelik, B. Sener, Prevalence and genetic diversity of *Staphylococcus aureus* small-colony variants in cystic fibrosis patients, Clin. Microbiol. Infect.: Off. Publ. Eur. Soc. Clin. Microbiol. Infect. Dis. 19 (2013) 77–84.
[56] S. Subramoni, D.T. Nguyen, P.A. Sokol, *Burkholderia cenocepacia* ShvR-regulated genes that influence colony morphology, biofilm formation, and virulence, Infect. Immun. 79 (2011) 2984–2997.