Species Identification of Coagulase-Negative Staphylococci: Genotyping is Superior to Phenotyping
Ruth N. Zadoks, Jeffrey L. Watts

To cite this version:

HAL Id: hal-00532480
https://hal.archives-ouvertes.fr/hal-00532480
Submitted on 4 Nov 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Accepted Manuscript

Title: Species Identification of Coagulase-Negative Staphylococci: Genotyping is Superior to Phenotyping

Authors: Ruth N. Zadoks, Jeffrey L. Watts

PII: S0378-1135(08)00360-X
Reference: VETMIC 4145

To appear in: VETMIC

Please cite this article as: Zadoks, R.N., Watts, J.L., Species Identification of Coagulase-Negative Staphylococci: Genotyping is Superior to Phenotyping, Veterinary Microbiology (2008), doi:10.1016/j.vetmic.2008.09.012

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.
Species Identification of Coagulase-Negative Staphylococci: Genotyping is Superior to Phenotyping

Ruth N. Zadoks *1) and Jeffrey L. Watts 2)

1) Quality Milk Production Services, Cornell University, Ithaca, NY

2) Pfizer Animal Health, Kalamazoo, MI

* Corresponding Author:

Ruth N. Zadoks

Current address: Moredun Research Institute

Pentlands Science Park

Bush Loan

Penicuik

EH26 0PZ

Scotland, UK

Phone: +44 (0)131 445 5111

E-mail: ruth.zadoks@moredun.ac.uk
Abstract

Coagulase-negative staphylococci (CNS) are isolated commonly from bovine milk and skin. Their impact on udder health and milk quality is debated. It has been suggested that sources and consequences of infection may differ between CNS species. Species-specific knowledge of the impact and epidemiology of CNS intramammary infections is necessary to evaluate whether species-specific infection control measures are feasible and economically justified. Accurate measurement of impact, sources, and transmission mechanisms requires accurate species level identification of CNS. Several phenotypic and genotypic methods for identification of CNS species are available. Many methods were developed for use in human medicine, and their ability to identify bovine CNS isolates varies. Typeability and accuracy of typing methods are affected by the distribution of CNS species and strains in different host species, and by the ability of test systems to incorporate information on new CNS species into their experimental design and reference database. Generally, typeability and accuracy of bovine CNS identification are higher for genotypic methods than for phenotypic methods. As reviewed in this paper, DNA sequence-based species identification of CNS is currently the most accurate species identification method available because it has the largest reference database, and because a universally meaningful quantitative measure of homology with known species is determined. Once sources, transmission mechanisms, and impact of different CNS species on cow health, productivity and milk quality have been identified through use of epidemiological data and accurate species identification methods, appropriate methods for routine use in research and diagnostic laboratories can be proposed.
Key Words: Genotyping, phenotyping, bovine, coagulase negative staphylococcus, mastitis
1. Introduction

The role of coagulase-negative staphylococci (CNS) as etiological agents of bovine mastitis has not been fully elucidated and previous studies in this area have yielded contradictory results. For example, some researchers regard CNS as an important cause of bovine mastitis (Pyörälä and Taponen, 2008), while others consider them minor pathogens with limited impact on milk quality and udder health (Schukken et al., 2008). Presence of CNS is associated with clinical mastitis and with somatic cell counts (SCC) that are, on average, higher than those in culture-negative quarters (Schepers et al., 1997; Kudinha and Simango, 2002). Increased SCC is generally associated with decreased milk production (Seegers et al., 2003) but subclinical intramammary infections by CNS have been associated with increased milk production (Wilson et al., 1997). By contrast, clinical CNS mastitis was linked to decreased milk production (Gröhn et al., 2004) and increased risk of culling (Gröhn et al., 2005). Mere detection of CNS in a milk sample was not associated with an increased risk of treatment or culling for mastitis (Reksen et al., 2006). In addition to debate about the impact and relevance of CNS infections, there is debate on whether or not specific CNS species are associated with the outcome of infection. A Finnish study showed an association between CNS species and severity of clinical symptoms (Honkanen-Buzalski et al., 1994), but a different study from the same country did not show such an association (Taponen et al., 2006). One CNS species, *Staphylococcus chromogenes*, is thought to protect the udder from intramammary infection (Matthews et al., 1990a; De Vliegher et al., 2003), whereas *Staphylococcus hyicus*, a closely related species, does not have this effect. CNS have been categorized
into human and animal-associated species (Watts and Owens, 1989). Human-associated CNS species, specifically *Staphylococcus epidermidis*, are thought to be more likely to invade and infect the udder than animal-associated species (Devriese and De Keyser, 1980; Watts and Owens, 1989). Prevalence of *Staphylococcus epidermidis* may be associated with herd management factors. Specifically, *S. epidermidis* is more common in herds that use linear dodecyl benzene sulphonyl acids for postmilking teat disinfection than in herds that use iodine (Hogan et al., 1987; Watts and Owens, 1989). Whether species specific transmission routes and control strategies exist for other CNS species is largely unknown. Because little species-specific information on control of CNS mastitis is available, identification to the group level, possibly supplemented with antimicrobial susceptibility testing, is currently sufficient for most therapeutic and management decisions. The defining characteristic of CNS as a group is the lack of coagulase expression, which is a phenotypic trait. To evaluate whether species-specific infection control measures are feasible and economically justified, species-specific knowledge of the impact and epidemiology of CNS infections is necessary. Accurate measurement of the impact, sources, transmission mechanisms and control options for individual CNS species requires accurate species level identification of CNS (Thorberg and Brändström, 2000; Heikens et al., 2005; Sivadon et al., 2005). In this contribution, merits of phenotypic and genotypic methods for CNS species identification are compared with special consideration of identification of CNS isolated from bovine milk.
2. The Species Concept

Interpretive criteria for the definition of bacterial genera and species are not consistent in the literature and may differ between species, genera and authors (Freney et al., 1999; Lan and Reeves, 2001; CLSI, 2007). In fact, the whole concept of what defines a bacterial species is a matter of debate (Lan and Reeves, 2001). Standards for description of new staphylococcal species were last defined by the Subcommittee on the taxonomy of staphylococci and streptococci of the International Committee on Systematic Bacteriology in 1999 and were based largely on phenotypic criteria (Freney et al., 1999). Nowadays, combinations of phenotypic and genotypic methods are used to define new species, such as biochemical profiling, gas chromatographic analysis of cellular fatty acids, ribotyping, sequencing of the 16S rRNA gene and sequencing of additional housekeeping genes (Becker et al., 2004; Carretto et al., 2005). For many species, there is only a single or a limited number of type strains, and their phenotype and genotype defines the species (Becker et al., 2004; Shah et al., 2007). The aim of this paper is not to discuss the definition of species, but the accurate designation of species names to clinical isolates. Given that species and type strains exist, our task as diagnosticians and scientists is to determine to which species the CNS isolates that we study belong.

3. Phenotypic Identification of CNS

Phenotypic identification methods are based on evaluation of the expression of genetically encoded characteristics by bacterial isolates. Phenotypic traits include
morphology, growth characteristics, ability to metabolize substrates, antimicrobial resistance, and other features that result from DNA-expression but that are not based on detection of the bacterial DNA itself. Over the years, many phenotypic methods have been developed for the identification of staphylococci in diagnostic laboratories. Methods include commercial test systems such as the API 20 Staph system (bioMérieux), API ID 32 Staph (bioMérieux), Staph-Zym (Rosco), the Vitek system (bioMérieux) and other combinations of biochemical tests, which may not be available in commercial formats (Bannerman et al., 1993; Devriese et al., 1994; Watts and Yancey, 1994; Ieven et al., 1995).

An inherent weakness of phenotypic methods is that there is variability in expression of phenotypic characteristics by isolates belonging to the same species (Bannerman et al., 1993; Ieven et al., 1995; Heikens et al., 2005). Furthermore, the interpretation of phenotypic tests can be subjective (Carretto et al., 2005). Variability in the expression and interpretation of phenotypic characteristics limits the reproducibility of tests, i.e. the ability to generate the same results every time the tests are used. In addition to reproducibility, the typeability and accuracy of phenotypic testing are imperfect. Typeability is the proportion of isolates that are assigned a type by a typing system (Struelens et al., 1996). An increase in the number of tests that is included in a system generally improves typeability. For example, a study of human CNS isolates with an API-system based on 20 biochemical reactions showed a typeability of only 37% (Carretto et al., 2005) whereas a system that included 32 reactions had a typeability of 85% (Maes et al., 1997).
Accuracy does not have a single standard definition. The concept can be interpreted in two major ways. First, accuracy can denote the level of certainty assigned by a test to its own results. The statement "isolate 13 was identified to *S. cohnii* by API Staph with 99.1% accuracy" (from Heir et al., 1999) would be an example of this interpretation. Second, accuracy can denote the level of agreement between a method and a reference method, i.e. the correctness of the identification. The statement "Isolate 13, identified to *S. cohnii* by API Staph with 99.1% accuracy was identified as *S. caseolyticus* by 16S rRNA gene sequence analysis " (based on Heir et al., 1999) shows that the 99.1% "accuracy" assigned to the phenotypic test score did not reflect correct species identification. In the same study, a second isolate was not identified by API Staph typing, and identified as *S. cohnii* by 16S rRNA gene sequencing and tDNA-ILP (tDNA intergenic spacer length polymorphism) (Heir et al., 1999). Results for the second isolate provide an example of limited typeability of the phenotypic method, but without incorrect or inaccurate species identification. The acceptable level of accuracy for phenotypic methods is subjective. In one study, 111 of 122 CNS isolates were identified by the Vitek GPI system, with 29% of those at more than 90% probability of accuracy (Lee and Park, 2001). The typeability of this system could be calculated as 111 of 122 = 91%, if all isolates for which a species name was generated were considered typeable, or as 32 of 122 = 26%, if only results with more than 90% probability of accuracy were considered acceptable for species identification. In a study on CNS isolates from ovine milk between 81.4% and 96.5% typeability was reported for three phenotypic species identification methods, but only 29% of isolates were identified in the same manner by all three methods (Burriel and Scott, 1998). This casts considerable doubt on the accuracy of the
species identification obtained with each method. In our opinion, high apparent accuracy resulting in incorrect species identification is a bigger concern than limited typeability. The latter alerts the user to the fact that the correct species identity was not determined by the test whereas the former gives a false sense of security. Other authors have also expressed concerns about false rather than ambiguous identification of CNS species by phenotypic systems (Sivadon et al., 2004).

Past identification errors or changes in taxonomy may contribute to incorrect species identification by phenotypic methods (Carretto et al., 2005). For example, \textit{S. felis} was named in 1989 (Igimi et al., 1989) and it is thought that many supposed \textit{S. simulans} isolates from cats that were identified before that time were really \textit{S. felis} isolates (Lilenbaum et al., 1999). It is unfair to penalize a system retroactively for misidentification of isolates prior to species reclassification, but once species reclassification occurs, test design, isolate identification and species distributions from studies preceding the reclassification may need to be reinterpreted. Until 1986, \textit{S. chromogenes} and \textit{S. hyicus} were not recognized as different species but seen as two subspecies, i.e. subspecies \textit{chromogenes} and subspecies \textit{hyicus}, of the species \textit{S. hyicus} (Hajek et al., 1986). In studies conducted after 1986, \textit{S. chromogenes} was often the most common CNS species found in milk (Table 1). Before 1986, these isolates would have been classified as \textit{S. hyicus}, which would be incorrect under current species definitions. Errors in identification are not limited to the species level, but may occur at the genus level. For example, some versions of the Vitek 2 system erroneously classified human clinical CNS isolates as \textit{Kocuria} species (Ben-Ami et al., 2005; Boudewijns et al., 2005). The distinction between \textit{Staphylococcus} and \textit{Kocuria} is not just of academic interest.
CNS isolates are viewed as pathogens and as an indication for treatment, while *Kocuria* species are considered contaminants that do not warrant treatment. Misclassification of *Staphylococcus* as *Kocuria* could result in withholding treatment from a patient, with potentially damaging consequences (Ben-Ami et al., 2005). The limitations of phenotypic methods are not unique to CNS. Similar concerns have been reported for other bacterial genera, such as *Enterococcus* and *Lactobacillus* in foods (Huys et al., 2006) and *Streptococcus* and *Enterococcus* isolates of animal origin (Watts and Yancey, 1994; Hudson et al., 2003; Loch et al., 2005).

Phenotypic methods are usually considered less expensive than genotypic methods. Whether or not this is true depends in part on turnover, which affects overhead costs and opportunities and needs for automation. In some clinical laboratories, phenotypic tests are used with such high frequency that an investment in automation of reading and interpretation of tests is profitable (Ieven et al., 1995). In other laboratories, test frequency may be so low that expiration of reagents and the costs of replacing them are a concern. Regardless of test volume, additional testing may be needed to obtain final results from phenotypic methods. This increases cost and turn-around-time of phenotyping testing, thereby narrowing or eliminating the cost and time differences between phenotypic and genotypic identification methods (Ieven et al., 1995; Thorberg and Brändström, 2000). When comparing the cost of phenotypic and genotypic methods, the costs of obtaining inaccurate results must also be considered.
4. Genotypic Identification of CNS

Genotypic methods use DNA as the basis for identification. Genotypic methods are used for identification to the species level and for strain typing, i.e. differentiation of isolates at the subspecies level. Genotypic methods used for identification of CNS species include amplified-fragment length polymorphism (AFLP) analysis (Taponen et al., 2006, 2007), ribotyping (Carretto et al., 2005), tDNA-ILP analysis (Maes et al., 1997; Heir et al., 1999; Lee and Park, 2001; Rossi et al., 2001; Stepanović et al., 2005), and DNA-sequencing (Heir et al., 1999; Sivadon et al., 2004; Heikens et al., 2005; Sivadon et al., 2005). Across bacterial genera the most common target for DNA-sequencing is the 16S rRNA gene (Lan and Reeves, 2001). Many CNS species are closely related, and 16S sequence based typing may not have sufficient discriminatory power to differentiate all of them (CLSI, 2007; Shah et al., 2007). Therefore, species identification systems based on the housekeeping genes cpn60 (chaperonin or heat-shock protein 60) (Kwok et al., 1999), dnaJ (heat-shock protein 40) (Shah et al., 2007), rpoB (beta subunit of RNA polymerase) (Drancourt and Raoult, 2002; Mellmann et al., 2006), sodA (superoxide dismutase A) (Poyart et al., 2001; Sivadon et al., 2004; Heikens et al., 2005; Sivadon et al., 2005), and tuf (elongation factor Tu) (Heikens et al., 2005) have been developed and implemented for CNS identification.

In general, genotypic methods have higher discriminatory power, reproducibility and typeability than phenotypic methods. For example, automated ribotyping identified 166 of 177 (94%) of CNS isolated from humans (Carretto et al., 2005). tDNA-ILP identified 157 of 161 (97.5%) S. sciuri group isolates from human, animal and
environmental sources (Stepanović et al., 2005), 162 of 163 (99%) of human clinical
CNS isolates from Belgium (Maes et al., 1997) and 114 of 122 (93%) of human clinical
isolates from Korea (Lee and Park, 2001). DNA-sequencing of CNS from bone and joint
infections identified 211 of 212 (99%) of isolates (Sivadon et al., 2005). An advantage of
DNA sequence-based methods is that they allow for recognition of previously unreported
sequences from novel species (Sivadon et al., 2004; CLSI, 2007).

Genotypic methods, like phenotypic methods, use cut-off values for acceptable
levels of similarity to identify bacterial species. For example, banding pattern similarity
of 90% or more is used as a cut-off value to consider isolates members of the same
bacterial species in automated ribotyping of CNS with restriction enzyme EcoRI
(Carretto et al., 2005). Categorization of banding patterns is often based on automated
analysis followed by visual inspection and manual correction, introducing some
subjectivity in the interpretation (Carretto et al., 2005) Automated data analysis is also
used for interpretation of tDNA-ILP results, but thresholds for species identification are
not defined clearly (Maes et al., 1997). Because tDNA-PCR is based on banding patterns
generated by PCR and separated by electrophoresis, results are somewhat susceptible to
PCR and electrophoresis conditions (Maes et al., 1997; Lee and Park, 2001). For analysis
of sequence data, interpretation criteria are gene-specific. For the highly conserved 16S
rRNA gene, 98% or 99% has been used as cut-off value (Nelson et al., 2003; Gill et al.,
2006; CLSI, 2007). Other housekeeping genes, such as sodA and tuf, are less conserved,
which allows them to be used in sequence-based strain typing methods (Zadoks et al.,
2005). For these genes, homology values of 97% or more are considered acceptable
(Heikens et al., 2005; Sivadon et al., 2005). Within-species heterogeneity of
housekeeping genes differs between bacterial species (Loch et al., 2005). Thus, both
within species-variability and between-species variability of sequence data may need to
be considered to decide on appropriate cut-off values for homology (Lan and Reeves,
2001). Some authors specifically include criteria for difference in sequence identity from
the next closest species, e.g. 5% or more, in guidelines for interpretation of DNA
sequence data of CNS species (Sivadon et al., 2005).

When genotypic methods were first developed, they tended to be more labor-
intensive than phenotypic methods, more expensive, or both. While some methods, such
as automated ribotyping, are still costly, other methods are not necessarily more
expensive than phenotypic methods. Affordability and feasibility of use of genotypic
methods differs between laboratory settings, just like affordability and feasibility of use
of phenotypic methods. Some laboratories have easy access to an automated RiboPrinter,
and other laboratories have in-house AFLP or tDNA-ILP facilities. DNA sequencing is a
highly portable method because material for sequencing can be shipped to specialized
laboratories by mail, and sequence data are routinely provided in electronic format for
remote downloading and analysis. There is no need for physical proximity of the
investigator or diagnostician to the sequencing facility. In one of our laboratories, species
identification of CNS isolates for diagnostic purposes is currently done by DNA
sequencing rather than phenotypic methods. Turnover of commercial phenotypic test kits
for identification in that laboratory is low. As a result, reagents tend to expire, and the
cost of replacing them increases the cost of phenotypic typing. By contrast, PCR and
DNA-sequencing are performed routinely so that reagents are fresh and CNS
identification fits into the workflow easily. Hands-on time per isolate is similar for
phenotypic and DNA sequence-based methods in this laboratory.

5. Databases

For interpretation of results from phenotypic or genotypic assays, comparison
with reference data is essential. Most phenotypic species identification methods were
developed for microbial isolates obtained from humans (Watts and Yancey, 1994).
Common CNS species in humans include, among others, *S. epidermidis*, *S. haemolyticus*,
*S. hominis*, *S. simulans*, *S. xylosus* and *S. capitis* (Maes et al., 1997; Lee and Park, 2001;
Carretto et al., 2005). Results of eight studies from seven countries and three continents
indicated that the most common CNS species isolated from bovine milk are *S.
chromogenes*, *S. hyicus*, *S. simulans*, *S. epidermidis* and *S. xylosus* (Table 1). *Staphylococcus chromogenes* and *S. hyicus*, the two most common CNS species from
bovine milk, were not recognized by an early version of the Staph-Zym system, while the
latter three species, all of which are also important in human medicine, were identified
correctly (Watts and Washburn, 1991). This is an example of the fact that most
commercial systems are developed for identification of human rather than animal
pathogens, and that the systems may not identify animal pathogens accurately. Early
versions of the Vitek and API Staph system also showed limited ability to identify *S.
chromogenes* and *S. hyicus* (Matthews et al., 1990b). Recent work in our laboratory
showed that API 20 Staph and BBL Crystal tests had low sensitivity and specificity in the
detection of *S. chromogenes* and *S. hyicus* isolates, respectively, when analyzing 82
isolates obtained in the USA from heifers around calving (unpublished data). In a second
dataset, encompassing 172 CNS isolates from lactating animals in The Netherlands,
StaphZym and API 20 Staph testing also showed limited typeability and accuracy. For
example, all *S. epidermidis* isolates were identified correctly but many isolates of *S.
*chromogenes* and *S. hyicus* were not (Sampimon et al., in preparation). Thorberg and
Brändström (2000) reported more favorable results for StaphZym analysis, i.e. 94%
accuracy across bacterial species. They suggest that cost may be the main impediment to
routine use of the method. Considering, however, that accuracy of detection of *S.
*chromogenes* was only 86%, even after additional tests were performed, and that *S.
*chromogenes* is the most common species isolated from bovine milk (Table 1), the value
of the test for routine diagnostics can be questioned.

Strain differences between bacterial isolates from different host species may
compound species identification problems (Watts and Yancey, 1994). For some Gram-
positive bacteria that are commonly found in humans and in bovine milk, specifically
*Streptococcus agalactiae* and *Staphylococcus aureus*, host-species specific groups of
strains have been identified (Smith et al., 2005; Sukhnanand et al., 2005). To our
knowledge, strain level comparisons of CNS isolates from humans and cattle have not
been reported. If host-species associated strains exist within CNS species, phenotypic
methods that were developed using human isolates may not identify bovine strains
correctly, even if the bacterial species is common to both host species.

The problem of a CNS-species distribution that differs from the one found in
humans is not unique to bovine CNS. CNS populations in other animal species and in
foods also differ from those in humans (Table 2). As a result, phenotypic methods that
were developed for identification of human CNS may also fail to identify CNS from other animal species accurately. Some CNS species are common to many hosts, including humans, e.g. *S. epidermidis*, *S. haemolyticus*, *S. saprophyticus*, *S. simulans*, and *S. xylosus*. Other CNS species, such as *S. caprae*, *S. chromogenes*, *S. felis*, *S. gallinarum* and *S. sciuri* are the most common species in small ruminants (Deinhofer and Pernthaner, 1995; Pengov 2001), cattle (Table 1), cats (Igimi et al., 1994; Lilienbaum et al., 1999), chickens (Awan and Matsumoto, 1998; Aarestrup et al., 2000) and treefrogs (Slaughter et al., 2001), respectively, while they are rare in other host species. Species that are rarely isolated from human clinical samples, e.g. *S. equorum*, *S. pasteuri* and *S. capitis* (relatively common in birds and horses, Table 2), are not included in the API 20 Staph database (Ieven et al., 1995; Carretto et al., 2005) or, in the case of *S. fleuretti* and *S. vitulinus*, in any commercial system (Stepanović et al., 2005). *Staphylococcus equorum* was the fifth-most common species isolated from milk in the Netherlands in the aforementioned study (Sampimon et al., in preparation). Because this species is not recognized by API 20 Staph testing, and because many studies of bovine CNS use API testing (Chaffer et al., 1999; Matthews et al., 1990b; Sampimon et al., 2007) the prevalence of *S. equorum* in milk may be underestimated. In contrast to the API 20 Staph test, the API Staph ID 32 test does have the ability to identify *S. equorum* from milk samples (Taponen et al., 2006).

Keeping databases up to date is a major challenge. In 1995, 31 *Staphylococcus* spp. had been named (Ieven et al., 1995). By 2003, this number had increased to 38 (Spergser et al., 2003). By 2007, the NCBI Taxonomy database (http://www.ncbi.nlm.nih.gov/) listed 43 named *Staphylococcus* spp. and more than 50
unnamed *Staphylococcus* spp. These numbers do not include subspecies within named or
unnamed species. Recently named CNS species include *Staphylococcus nepalensis,*
which was first isolated from a goat in the Himalayas and identified as a new species in
2003 (Spergser et al., 2003), and *Staphylococcus fleuretti,* which was first named in 2000
(Vernozy-Rozand et al., 2000). Among CNS isolates from bovine milk that were
characterized in our laboratory, both *S. nepalensis* and *S. fleuretti* have been identified
using DNA sequencing and comparison with on-line databases. Phenotypic systems for
routine diagnostic use cannot be updated every time a new bacterial species is identified
or every time strain level differences between isolates from different host species are
recognized. By contrast, it is extremely easy to add sequences from new strains or species
to a reference DNA database such as GenBank shortly after detection. The availability of
reference data affects the typeability and accuracy of phenotypic as well as genotypic
methods. Like phenotypic methods, genotypic methods that have been used for
identification of human CNS but not bovine CNS may suffer from limited typeability
when bovine isolates are first characterized. For example, AFLP failed to identify 19 of
99 isolates (19%) upon its first use with a reference database of 39 staphylococcal species
and subspecies (Taponen et al., 2006). In a subsequent study using a reference database
of 48 species and subspecies, including *S. equorum* and *S. fleuretti,* only 11 of 120
isolates (9%) could not be identified (Taponen et al., 2007). Although isolate collections
differed between the two studies (Taponen et al., 2006, 2007), it seems reasonable to
infer that typeability increased when the reference database of the system was expanded.
Initial comparison of DNA sequence data with tDNA-PCR data showed only 75%
typeability of bovine CNS by tDNA-PCR and 73% agreement with DNA sequence data.
With use of sequence-based species identification, previously unidentified tDNA-PCR patterns could be named and added to the tDNA-PCR database to improve typeability (Supré et al., in preparation).

Because the use of DNA-sequencing for CNS identification is relatively new, interpretive criteria are still under development (CLSI, 2007). Some genes are so conserved that differentiation of species or subspecies based on sequence data is not even possible. For example, 16S rRNA gene sequencing failed to distinguish *S. caprae* from *S. capitis*, whilst automated ribotyping was able to differentiate the two species (Carretto et al., 2005). Interpretation of sequence data for genes that are less conserved than 16S can be a challenge too. When we first started to use *rpoB* sequencing for identification of bovine CNS in our laboratory in 2005, many *S. hyicus* isolates could not be identified with certainty based on *rpoB* data alone, because the homology between *rpoB* sequence data available in GenBank and those obtained from milk isolates was 94%, which is below the 97% homology criterion that was suggested at the time for housekeeping genes other than 16S (Heikens et al., 2005; Sivadon et al., 2005). The next best match in the *rpoB* database showed 89% homology, which did meet the criterion of 5% difference to the next bacterial species (Sivadon et al., 2004). Sequencing of additional housekeeping genes, i.e. the 16S rRNA and *cpn*60 genes, confirmed the isolates as *S. hyicus* with more than 99% homology. It appears that within the species *S. hyicus*, considerable sequence diversity exists within the *rpoB* gene. Once a larger variety of *rpoB* alleles is added to the on-line database, identification of *S. hyicus* based on this gene should no longer be a problem. In the mean time, Mellmann et al. (2006) have proposed to use 94% homology as the cut-off value for *rpoB* homology in CNS species, underscoring that there is
considerable within-species heterogeneity in the DNA sequence of this gene. By this criterion, our original *rpoB* sequence data would have been sufficient to categorize all potential *S. hyicus* isolates as such. Whether the rule that the difference in sequence homology to the next best match should be 5% (Sivadon et al., 2004) should be maintained is doubtful. For some species, e.g. *S. fleuretti* vs. *S. pulveri* or *S. hæmolyticus* vs. *S. croceolyticus*, the difference in DNA sequence homology is routinely 4%, exemplifying that the difference between within-species and between-species homology may be more relevant than a single pre-defined cut-off value. As the uptake by laboratories of DNA sequencing for microbial identification increases, more reference data and standard protocols for generating and interpreting of sequence data will become available (CLSI, 2007).

One undeniable advantage of DNA sequence-based methods over all phenotypic and genotypic methods is that it provides a quantitative measure, down to the last base pair of the genetic code, of the certainty with which an isolate has been identified (CLSI, 2007). For other genotypic methods, such as automated ribotyping and tDNA-ILP, some of the limitations of phenotypic methods apply, i.e. similarity coefficients can be calculated but there is no universally meaningful quantitative measure of the genetic relatedness of isolates.

6. Discussion

Efforts are underway to compare species identification of bovine CNS by genotypic and phenotypic methods, as documented in references cited in this paper, or in
manuscripts that were still in preparation at the time this paper was written. Using DNA-sequencing, over 99% of CNS isolates from bovine milk can be identified. So far, other genotypic methods and phenotypic methods have lower typeability and accuracy, which may manifest in missing, incorrect, or ambiguous results. As databases grow, other methods may become more accurate, although database updates may not keep pace with the increase in number of described *Staphylococcus* species and subspecies.

For diagnostic work, the choice of typing methods should be determined by a number of considerations, such as the goal of isolate identification, speed, ease of use, cost, availability of equipment and trained personnel, etc. In the case of CNS identification in the context of mastitis control, the best method would be a fast, simple and cheap method that provides a relevant level of differentiation. The amount of information regarding clinical relevance, treatment or management of CNS mastitis that is based on accurate species level identification is limited. Until proven otherwise, CNS may be the most relevant level of identification for mastitis diagnostics, and this level is achieved through phenotypic methods. If CNS species of specific importance to udder health and mastitis control are identified, simplified phenotypic or genotypic methods targeting this subset of CNS may be of value for routine diagnostics. Simplified phenotypic diagnostic methods have been developed for identification of clinically significant CNS species in hospitals (Ieven et al., 1995). Such simplified methods can be of great utility if they are rapid and inexpensive, and if they do not misidentify isolates belonging to species that are not covered by the simplified method. For example, in the hypothetical situation that *S. chromogenes* and *S. epidermidis* were shown to have positive (Matthews et al., 1990a; De Vliegher et al., 2003) and negative (Devriese and De
Keyser, 1980; Watts and Owens, 1989) impact on udder health, respectively, whereas other CNS species are merely innocent bystanders, it would suffice to have a system that classified CNS isolates as *S. chromogenes*, *S. epidermidis*, or "other CNS species". This would be similar to existing diagnostics for streptococci from bovine milk, which are commonly differentiated into *Strep. agalactiae*, *Strep. dysgalactiae*, or *Strep. iberis* with the remaining group comprised of other species of streptococci and enterococci. Unidentified isolates from rest groups can be identified by additional testing as necessary. Simplified schemes for routine diagnostics could be phenotypic or genotypic. A simplified phenotypic scheme for identification of CNS from bovine mastitis has already been proposed (Devriese et al., 1994). If clinically relevant, a simplified genotypic scheme, such as a multiplex PCR for a limited number of clinically relevant CNS species, could be developed, as has been done for the most clinically relevant *Streptococcus* spp. (Phuektes et al., 2001). In light of the discovery of new species, such as *S. equorum*, which was in the top-5 of most common species in our laboratory, and in changes in identification methods, existing simplified diagnostic schemes may need to be reevaluated and updated. When evaluating simplified identification schemes, it is important to assess sensitivity, i.e. the ability to recognize known members of the species, and specificity, i.e. the ability to exclude isolates that are not members of the species. The latter is not always done (Devriese et al., 2002). As for any other diagnostic test, the positive and negative predictive value of a test result will not only depend on the sensitivity and specificity of the test, but also on the composition of the bacterial population. This composition may differ significantly between dairy cattle, humans and...
other host species, and a test system that is useful in one diagnostic or research setting may not be as useful in a different setting.

7. Conclusion and Outlook

DNA sequence-based species identification of CNS is currently the most accurate species identification method available because it has the largest reference database, and because a universally meaningful quantitative measure of homology with known species is determined. DNA sequence-based species identification could therefore be considered the Gold standard and should be used as the reference identification methodology. If a diagnostic laboratory does not have access to a reference method, or a method that has been validated through comparison with a reference method, characterization of CNS isolates is best limited to actual observations, such as coagulase reaction, novobiocin resistance, etc. Reporting as "coagulase-negative Staphylococcus species" may be more appropriate than reporting of more detailed but potentially inaccurate results. At present, species-specific recommendations for management and control of CNS mastitis in dairy herds have not been formulated, and identification of CNS as such will suffice for routine diagnostics. In research, genotypic methods for species identification are to be preferred over phenotypic methods. Once the source, transmission mechanisms, and impact of different CNS species on cow health, productivity and milk quality have been identified through use of epidemiological data and accurate species identification methods, appropriate identification methods for routine use in research and diagnostic laboratories can be proposed.
Acknowledgements

The work of R.N. Zadoks was supported in part by the Multi-State Mastitis Research Project USDA CSREES NE-1028 (formerly NE-1009).

Conflict of Interest

Neither one of the authors (R.N. Zadoks, J. L. Watts) has a financial or personal relationship with other people or organizations that could inappropriately influence or bias the paper entitled "Species Identification of Coagulase-Negative Staphylococci: Genotyping is Superior to Phenotyping". J. L. Watts is an employee of Pfizer Animal Health, Kalamazoo, MI.
Table 1

Distribution of coagulase-negative staphylococcal species (in % of isolates tested) in bovine milk. Based on phenotypic and genotypic methods as specified in Matthews et al., 1990b; Watts and Washburn, 1991; Devriese et al., 1994; Waage et al., 1999; Thorberg and Brändström, 2000; Kudinha and Simango, 2002; Taponen et al., 2006; Sampimon et al., 2007.

<table>
<thead>
<tr>
<th>Species</th>
<th>Belgium</th>
<th>Finland</th>
<th>NL</th>
<th>Norway</th>
<th>Sweden</th>
<th>USA</th>
<th>USA</th>
<th>Zim</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TN</td>
<td>LA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S. auricularis</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>S. capitis</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>S. caseolyticus</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>S. chromogenes</td>
<td>12</td>
<td>27</td>
<td>42</td>
<td>15</td>
<td>17</td>
<td>40</td>
<td>11</td>
<td>24</td>
<td>24</td>
</tr>
<tr>
<td>S. cohnii</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>S. epidermidis</td>
<td>9</td>
<td>5</td>
<td>7</td>
<td>0</td>
<td>13</td>
<td>5</td>
<td>12</td>
<td>23</td>
<td>9</td>
</tr>
<tr>
<td>S. equorum</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>S. fleuretti</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>S. haemolyticus</td>
<td>0</td>
<td>3</td>
<td>7</td>
<td>2</td>
<td>8</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>S. hominis</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>7</td>
<td>18</td>
<td>4</td>
</tr>
<tr>
<td>S. hyicus</td>
<td>5</td>
<td>2</td>
<td>6</td>
<td>15</td>
<td>12</td>
<td>12</td>
<td>35</td>
<td>8</td>
<td>12</td>
</tr>
<tr>
<td>S. intermedius</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>S. kloosii</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>S. lentus</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>S. muscae</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>S. saprophyticus</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>S. sciuiri</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>6</td>
<td>6</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>S. simulans</td>
<td>34</td>
<td>36</td>
<td>10</td>
<td>54</td>
<td>16</td>
<td>13</td>
<td>3</td>
<td>1</td>
<td>22</td>
</tr>
<tr>
<td>S. warneri</td>
<td>25</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>7</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>S. xylosus</td>
<td>12</td>
<td>0</td>
<td>16</td>
<td>3</td>
<td>13</td>
<td>9</td>
<td>17</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>unknown</td>
<td>0</td>
<td>19</td>
<td>0</td>
<td>8</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>7</td>
<td>5</td>
</tr>
</tbody>
</table>

1) NL = the Netherlands, USA = United States of America, Zim = Zimbabwe.

2) TN = Tennessee, LA = Louisiana. Two states were included to reflect some of the geographic diversity within the USA.

3) Total calculated as weighted average of results for each studied, where weight equals the number of isolates per study.
Table 2

Distribution of coagulase-negative staphylococcal species from various host species and food. Based on phenotypic and genotypic methods as specified in Cox et al., 1988; Igimi et al., 1994; Deinhofer and Pernthaner, 1995; Madsen and Christensen, 1995; Maes et al., 1997; Awan and Matsumoto, 1998; Lee et al., 1998; Lilenbaum et al., 1999; Aarestrup et al., 2000; Lilenbaum et al., 2000; Pengov, 2001; Slaughter et al., 2001; Carretto et al., 2005 and in Table 1.

<table>
<thead>
<tr>
<th>CNS species</th>
<th>Avian</th>
<th>Bovine</th>
<th>Canine</th>
<th>Caprine</th>
<th>Equine</th>
<th>Feline</th>
<th>Human</th>
<th>Ovine</th>
<th>Sausage</th>
<th>Treefrog</th>
</tr>
</thead>
<tbody>
<tr>
<td>S. capitis</td>
<td>X</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>X</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>S. caprae</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>X</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>X</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>S. camosus</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>X</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>S. chromogenes</td>
<td>-</td>
<td>X</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>S. cohnii</td>
<td>X</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>S. epidermidis</td>
<td>-</td>
<td>-</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>S. felis</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>X</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>S. gallinarum</td>
<td>X</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>S. haemolyticus</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>S. hyicus</td>
<td>X</td>
<td>X</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>X</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>S. lentus</td>
<td>X</td>
<td>-</td>
<td>X</td>
<td>X</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>S. saprophyticus</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>X</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>S. sciuri</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>X</td>
</tr>
<tr>
<td>S. simulans</td>
<td>X</td>
<td>X</td>
<td>-</td>
<td>X</td>
<td>-</td>
<td>X</td>
<td>X</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>S. xylosus</td>
<td>X</td>
<td>-</td>
<td>X</td>
<td>-</td>
<td>X</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>
References


Gill, J. J., Sabour, P. M., Gong, J., Yu, H., Leslie, K. E., Griffiths, M. W. 2006. Characterization of bacterial populations recovered from the teat canals of...


Lan, R., Reeves, P. R. 2001. When does a clone deserve a name? A perspective on bacterial species based on population genetics. Trends Microbiol. 9, 419-424.


