TITLE PAGE

Prediction of whole-genome DNA G+C content within the genus *Aeromonas* based on housekeeping gene sequences

J. Gaspar Lorén, Maribel Farfán, David Miñana-Galbis, and M. Carmen Fusté*

Departament de Microbiologia i Parasitologia Sanitàries, Facultat de Farmàcia, Universitat de Barcelona, Barcelona, Spain.

Running title:

Prediction of G+C content in *Aeromonas* species

*Corresponding author:

Dr. M.Carmen Fusté

Mailing address: Departament de Microbiologia i Parasitologia Sanitàries, Facultat de Farmàcia, Universitat de Barcelona, Avda. Joan XXIII s/n, 08028. Barcelona, Spain.

Phone: 34 93 402 44 97.

Fax: 34 93 402 44 98.

E-mail: mcfuste@ub.edu

Footnote:

Nucleotide sequences obtained in this work are deposited in the GenBank database under accession numbers: FJ936120, and FJ936121 (for *cpn60*); FJ936122, FJ936124, FJ936129, and FJ936130 (for *dnaJ*); FJ936131 (for *rpoB*); FJ936132, and FJ936133 (for *rpoD*).

ABSTRACT

Different methods are available to determine the G+C content (e.g. thermal denaturation temperature or High Performance Liquid Chromatography, HPLC), but obtained values may differ significantly between strains as well as between laboratories. Recently, several authors [7, 14] demonstrated that the genomic DNA G+C content of prokaryotes can be reliably estimated from one or several protein coding gene nucleotide sequences. Few G+C content values have been published for the *Aeromonas* species described, and the data when available are often incomplete or only provide a range of values. Our aim in this current work was twofold. First, we determined the genomic G+C content of the type or reference strains of all species and subspecies of the genus *Aeromonas* with a traditional experimental method in the same laboratory. Second, we wanted to see if the sequence-based method to estimate the G+C content described by Fournier *et al.* [7] could be applied to determine the G+C content of the different species of *Aeromonas* from sequences of the genes used in taxonomy or phylogeny in this genus.

Keywords:

G+C content

Aeromonas

Scope:

Systematics

INTRODUCTION

The DNA base composition is one of the most straightforward genomic characteristics to measure, and has been determined in thousands of bacteria, in which the genomic guanine plus cytosine content ranges from 25 to 77 mol% [8]. Many evolutionary mechanisms have been proposed to explain this G+C content diversity among bacteria, but most authors agree that the genomic G+C content of a species is set by a balance between selective constraints at the level of codons and amino acids and directional mutational pressure at the nucleotide level [33, 8].

The determination of the base composition of deoxyribonucleic acid is a key parameter in prokaryotic genomes that is usually used in taxonomic classification. The current recommendation for the description of a novel bacterial species is based on a polyphasic approach, including the determination of the G+C content as well as other characteristics such as DNA-DNA relatedness and phylogenetic classification [32].

Several different methods are available to determine the G+C content (e.g. thermal denaturation temperature or High Performance Liquid Chromatography, HPLC), but obtained values may differ significantly between strains as well as between laboratories. The thermal denaturation temperature (T_m) method is one of the most common techniques for determining this value, based on monitoring the increase of absorbance at 260 nm during DNA denaturation [18]. The T_m of DNA is influenced by the ionic strength of the DNA solution, and thus the value is difficult to reproduce from one laboratory to another. To minimize experimental errors, a reference DNA is used as a standard, and the G+C content is calculated by a formula reported by Mandel *et al.* [17]. However, this formula can not be applied to prokaryotes that have an extremely high or low G+C content, as the resulting value differs from those obtained by HPLC [5, 34]. For all these methodological reasons, a variation of up to 5% is generally accepted in the G+C content value within a single species [9]. Currently, the thermal denaturation temperature method has almost been substituted by the HPLC technique [23]. The HPLC method is more rapid and sensitive, but has disadvantages in cost and methodological complexity.

Recently, several authors [7, 14] demonstrated that the genomic DNA G+C content of prokaryotes can be reliably estimated from one or several protein coding gene nucleotide sequences. So far, this methodological approach has been applied to several phylogenetic distant bacteria [7] and strains belonging to different genera of the family *Pasteurellaceae* [14]. These authors have concluded that the sequence–based method is congruent with data obtained from conventional methods, reproducible, rapid and less labour-intensive.

In this study, we developed a method to predict the genomic G+C content in the genus *Aeromonas* at the interspecific level. The genus *Aeromonas* Stanier 1943 comprises Gram-negative, non-sporing, oxidase- and catalase-positive, facultatively anaerobic bacilli that are resistant to vibriostatic agent O/129 and are generally motile by means of a polar flagellum. They reduce nitrate to nitrite and do not require NaCl for growth [1, 19]. Taxonomically, this genus belongs to the family *Aeromonadaceae* and seems to form a monophyletic group in the γ-subgroup of the class *Proteobacteria* [19]. They are often associated with aquatic animals and frequently isolated from foods. There is strong evidence for the role of aeromonads as aetiological agents of a variety of infections in ectothermic animals (fish, frogs, turtles and snails). During the last 20 years the genus *Aeromonas* has been increasingly recognized as an agent of disease in humans, and associated with a variety of clinical manifestations. However, the correlation between species and disease remains to be elucidated and requires additional information about the taxonomy of these ubiquitous bacteria [19, 6].

The classification of the genus *Aeromonas* remains complex from a taxonomical point of view due to the continuous description of novel species, the rearrangement of strains and species described thus far, and the discrepancies observed in different DNA–DNA hybridization studies [10, 11, 13, 20, 25]. Recent studies based on the partial sequences of *cpn60*, *dnaJ*, *gyrB*, *rpoB,* and *rpoD* genes have shown that the use of several housekeeping genes is an effective approach to the phylogeny and taxonomic identification of *Aeromonas* species [31, 15, 29, 27, 26].

Our aim in this current work was twofold. Few G+C content values have been published for the *Aeromonas* species described, and the data when available are often incomplete or only provide a range of values. Our first objective was thus to determine the genomic G+C content of the type or reference strains of all species and subspecies of the genus *Aeromonas* with a traditional experimental method in the same laboratory. Secondly, we wanted to see if the sequence-based method to estimate the G+C content described by Fournier *et al.* [7] could be applied to determine the G+C content of the different species of *Aeromonas* from sequences of the genes used in taxonomy or phylogeny in this genus.

METHODS

Bacterial strains

We have analyzed a collection of 31 strains belonging to the genus *Aeromonas* (Table 1). This collection includes all the species and subspecies recognized up to June 2009 [29, 26], some strains considered synonyms, such as *A. ichthiosmia*/*A.veronii* [11], *A. enteropelogenes*/*A. trota* [12] and

A.culicicola/*A.veronii* [13], and also reclassified strains, such as *Aeromonas* DNA hybridization group 11 in *A. encheleia* [10], and *A. aquariorum,* which has been recently reclassified as *A. hydrophila* subsp. *dhakensis* [22]. We excluded *A. sharmana* from this study because it has been proven that it does not belong to this genus [21], and also the very recently accepted new strains, such as *A. fluvialis* [2], *A. piscicola* [4], *A. taiwanensis* and *A. sanarelli* [3] have not been considered.

DNA G+C content determination

The G+C content of genomic DNA was determined experimentally by the HPLC (high performance liquid chromatography) technique [23] at the BCCM^{TM}/LMG (Belgian Co-ordinated Collections of Microorganisms / Laboratorium voor Microbiologie) Identification Service of University of Gent (Belgium). The G+C values are expressed as percentages (mol%).

Gene sequences

We selected five conserved genes widely used in taxonomic classification and phylogeny of *Aeromonas* (*cpn60*, *dnaJ*, *gyrB*, *rpoB* and *rpoD*). The nucleotide sequences of these genes were obtained from the GenBank database for the strains used in this work. Nine sequences not included in the database were determined in our laboratory according to the methods previously described (*cpn60*, *dnaJ*, *rpoB*, *rpoD*) [31, 15, 27, 26]. All GenBank accession numbers from the nucleotide sequences used in this study are indicated in Table 1.

Statistical analysis

All statistical analysis was carried out using R software [28] and EXCEL spreadsheet (Microsoft). The statistical significance of the regression analysis between the experimental genomic G+C content and the G+C content calculated from the sequences of the *cpn60*, *dnaJ*, *gyrB*, *rpoB* and *rpoD* genes was determined using the *t*-test [$t = r\sqrt{(n-2)}/\sqrt{(1-r^2)}$], where r is the Pearson's correlation coefficient, r^2 is the coefficient of determination and n represents the number of species analyzed [16]. As a measure of the goodness of each regression model we used the coefficient of determination (r^2) and Akaike's information criterion (AIC). AIC was obtained using the *stats* package for R software and calculated as AIC = n ln(RSS/n) + 2p + n ln(2π), where n is the number of observations (31), *p* represents the number of parameters in the model (2) and RSS the residual sum of squares of the linear regression model. Given a data set, several competing models may be ranked according to their AIC, with the one having the lowest AIC being the best [16]. Observed differences were considered significant when $P < 0.05$.

RESULTS AND DISCUSSION

Experimental determination of G+C content

At present, the DNA G+C content has only been reported in a few species and subspecies of the genus *Aeromonas* (Table 2). In this study we experimentally determined the genomic G+C content of 31 type and reference strains of the species and subspecies of *Aeromonas* (Table 2). The variation in the G+C content for this genus was 5.3%, ranging from a minimum of 57.4% (*A. sobria*) to a maximum of 62.7% (*A. encheleia*), which is in agreement with those published previously (57-63% [19]). The difference in DNA G+C content obtained falls within the accepted values (<10 mol%) for microorganisms belonging to the same genus [9].

G+C content from housekeeping gene sequences

As reported by Fournier *et al.* [7] and Kuhnert & Korczak [14] the DNA G+C genomic content can be accurately estimated from the sequences of one or more protein codifying genes. We determined the G+C content of each strain analyzed from the *cpn60*, *dnaJ*, *gyrB*, *rpoB*, and *rpoD* gene sequences. The range, extreme values and the median of G+C content calculated from these sequences compared with the values obtained experimentally are shown in Figure 1.

Correlation between experimental and sequence gene methods

We performed a regression analysis between the experimental DNA G+C and the G+C content calculated from the sequence of each of the aforementioned five genes. The regression equations and the Pearson´s correlation coefficients (*r*) as well as their significance are shown in Table 3. Two of the five selected genes, *dnaJ* and *rpoB*, were later excluded from this study because of their low significance (*r* and AIC values). The average values obtained from the sequences of the three remaining genes (*cpn60*, *gyrB* and *rpoD*) were used to perform a regression analysis with the G+C content experimentally determined (Table 3). As the sequences of the three chosen genes differed in length, we weighed their average G+C content values with the mean length of the sequences (Table 3). However, the differences between the weighed average and the regression analysis performed with the simple mean were minimal (data not shown). The scatter plot, regression line as well as the regression equation and the coefficient of determination are shown in Figure 2. The value of the coefficient of determination obtained (r^2= 0.8326) is reasonably good, and suggests that this method is a reliable way of estimating the G+C content of *Aeromonas* species. The results obtained using this regression equation (3 genes) for each of the analyzed strains are shown in Table 2. The difference between the experimentally determined and the predicted values did not exceed 3% (Table 2), thereby being within the range of variation observed in G+C content determination with conventional methods [9].

As a way of checking the reliability of our approach, we inferred the G+C content of four strains of *A. molluscorum* not included in the previous analysis, by using the regression equation shown in Table 3. Those strains were chosen because we had previously experimentally determined their G+C content. Similarly, we also calculated the G+C content of the two *Aeromonas* species (*A. hydrophila* ATCC 7966^T and *A. salmonicida* A449) whose genomes have been sequenced. The results obtained were very precise and the absolute differences did not exceed 1% (Table 4).

In order to examine the intraspecies variation, we calculated the G+C content from the sequences of *cpn60*, *gyrB* and *rpoD* genes in a collection of 50 strains belonging to *A. bestiarum*, *A. hydrophila*, *A. molluscorum* and *A. salmonicida*. As seen in Table 5 all the standard error values ranged between 0.1 and 0.2 mol%, except in the case of *cpn60* for *A. molluscorum* (0.4 mol%). The higher variation observed in *A. molluscorum* is due to anomalous value (60.7 mol%) obtained from the strain 849T. Despite this rather high value, all the data obtained are well below those obtained for this genus interspecifically.

Selection of *cpn60*

Since sequence determination of three genes might sometimes be cumbersome, we have investigated if one of these genes alone might be representative of the whole. Recently, we have demonstrated that *cpn60*, whose sequencing is simple and rapid, is a good genetic marker for the *Aeromonas* species identification [26]. In order to investigate if the *cpn60* gene could be suitable candidate, we have performed a regression analysis of the G+C content calculated using *cpn60* sequences versus the values calculated using the weighed average of *cpn60*, *gyrB* and *rpoD*. The scatter plot, regression line as well as the regression equation and the coefficient of determination of this analysis are shown in Figure 3. In addition, data of regression analysis of G+C content calculated from *cpn60* sequences versus G+C content experimentally determined are indicated in Table 3. The value of the coefficient of determination obtained (r^2= 0.8181) indicated that there is a good correlation between the *cpn60* G+C content values and those obtained from the three genes, and allow us to suggest that the cpn60 sequences might be representative of all the genes studied.

Table 2 shows the predicted G+C content using only *cpn60* sequences for all the strains analyzed in this study. A mean difference of 0.66 mol% ± 0.53 was observed, which is only slightly higher than that obtained when using the regression model for all the three genes. These values are also within the range of variation observed in G+C content determination with conventional methods. Table 4 also shows the predicted values obtained with the same strains but using the regression equation of *cpn60*. The results were very similar to those obtained using the regression equation of the three genes.

In summary, in this study we have demonstrated that the genomic DNA G+C content of different species or subspecies of the genus *Aeromonas* can be estimated reliably from gene sequences. The results confirming those previously obtained by other authors [7, 14] with higher taxa. It is especially interesting that we were able to match the accuracy of experimental methods when determining the G+C content of the analyzed strains from the rapidly sequenced *cpn60.*

ACKNOWLEDGMENTS

This research was supported by projects CGL2004-03385/BOS and CGL2008-03281/BOS from the Ministerio de Educación y Ciencia, Spain.

REFERENCES

[1] Abbott, S.L., Cheung, W.K., Janda, J.M. (2003) The genus *Aeromonas*: biochemical characteristics, atypical reactions, and phenotypic identification schemes. J. Clin. Microbiol. 41, 2348-2357.

[2] Alperi, A., Martínez-Murcia, A.J., Monera, A., Saavedra, M.J., Figueras, M.J. (in press). *Aeromonas fluvialis* sp. nov., isolated from a Spanish river. Int. J. Syst. Evol. Microbiol. doi:10.1099/ijs.0.011643-0.

[3] Alperi, A., Martínez-Murcia, A.J., Ko, W.C., Monera, A., Saavedra, M.J., Figueras, M.J. (in press). *Aeromonas taiwanensis* sp. nov. and *Aeromonas sanarelli* sp. nov., two new clinical species from Taiwan. Int. J. Syst. Evol. Microbiol. doi:10.1099/ijs.0.014621-0.

[4] Beaz-Hidalgo, R., Alperi, A., Figueras, M.J., Romalde, J.L. (2009) *Aeromonas piscicola* sp. nov., isolated from diseased fish. Syst. Appl. Microbiol. 32, 471-479.

[5] Ezaki, T., Saidi, S.M., Liu, S.L., Hashimoto, Y., Yamamoto, H., Yabuuchi, E. (1990) Rapid procedure to determine the DNA base composition from small amounts of gram-positive bacteria. FEMS Microbiol. Lett. 55, 127-130.

[6] Figueras, M.J. (2005) Clinical relevance of *Aeromonas*. Rev. Med. Microbiol. 16, 145-153.

[7] Fournier, P.E., Suhre, K., Fournous, G., Raoult, D. (2006) Estimation of prokaryote genomic DNA G+C content by sequencing universally conserved genes. Int. J. Syst. Evol. Microbiol. 56, 1025-1029.

[8] Galtier, N., Lobry, J.R. (1997) Relationships between genomic G+C content, RNA secondary structures, and optimal growth temperature in prokaryotes. J. Mol. Evol. 44, 632-636.

[9] Goodfellow, M., Manfio, G.P., Chun, J. (1997) Towards a practical species concept for cultivable bacteria. In *Species: the Units of Biodiversity*, pp. 25-29. Edited by M.F. Clarridge, H.A. Dawah, M.R. Wilson. London: Chapman and Hall.

[10] Huys, G., Kämpfer, P., Altwegg, M., Coopman, R., Janssen, P., Gillis, M., Kersters, K. (1997) Inclusion of *Aeromonas* DNA hybridization group 11 in *Aeromonas encheleia* and extended descriptions of the species *Aeromonas eucrenophila* and *A. encheleia.* Int. J. Syst. Bacteriol. 47, 1157-1164.

[11] Huys, G., Kämpfer, P., Swings, J. (2001) New DNA-DNA hybridization and phenotypic data on the species *Aeromonas ichthiosmia* and *Aeromonas allosaccharophila*: *A. ichthiosmia* Schubert *et al.* 1990 is a later synonym of *A. veronii* Hickman-Brenner *et al.* 1987. Syst. Appl. Microbiol. 24, 177-182.

[12] Huys, G., Denys, R., Swings, J. (2002) DNA–DNA reassociation and phenotypic data indicate synonymy between *Aeromonas enteropelogenes* Schubert *et al.* 1990 and *Aeromonas trota* Carnahan *et al.* 1991. Int. J. Syst. Evol. Microbiol. 52, 1969–1972.

[13] Huys, G., Cnockaert, M., Swings, J. (2005) *Aeromonas culicicola* Pidiyar *et al.* 2002 is a later subjective synonym of *Aeromonas veronii* Hickman-Brenner *et al.* 1987. Syst. Appl. Microbiol. 28, 604–609.

[14] Kuhnert, P., Korczak, B.M. (2006) Prediction of whole-genome DNA-DNA similarity, determination of G+C content and phylogenetic analysis within the family *Pasteurellaceae* by multilocus sequence analysis (MLSA). Microbiology 152, 2537-2548.

[15] Küpfer, M., Kuhnert, P., Korczak, B.M., Peduzzi, R., Demarta, A. (2006) Genetic relationships of *Aeromonas* strains inferred from 16S rRNA, *gyrB* and *rpoB* gene sequences. Int. J. Syst. Evol. Microbiol. 56, 2743-2751.

[16] Maindonald, J., Braun, W.J. (2007) Data analysis and graphics using R: an example-based approach. 2nd ed. Cambridge; New York: Cambridge University Press.

[17] Mandel, M., Igambi, L., Bergendahl, J., Dodson, M.L. Jr, Scheltgen, E. (1970) Correlation of melting temperature and cesium chloride buoyant density of bacterial deoxyribonucleic acid. J. Bacteriol. 101, 333-338.

[18] Marmur, J., Doty, P. (1962) Determination of the base composition of deoxyribonucleic acid from its thermal denaturation temperature. J. Mol. Biol. 5, 109-118.

[19] Martin-Carnahan, A., Joseph, S.W. (2005) Genus I. *Aeromonas Stanier 1943, 213*^{AL}. In *Bergey´s Manual of Systematic Bacteriology*, 2nd ed, vol 2, part B, pp. 557-578. Edited by G.M. Garrity, D.J. Brenner, N.R. Krieg, J.T. Staley. New York: Springer.

[20] Martínez-Murcia, A.J. (1999) Phylogenetic positions of *Aeromonas encheleia*, *Aeromonas popoffii*, *Aeromonas* DNA hybridization group 11 and *Aeromonas* group 501. Int. J. Syst. Bacteriol. 49, 1403–1408.

[21] Martínez-Murcia, A.J., Figueras, M.J., Saavedra, M.J., Stackebrandt, E. (2007) The recently proponed species *Aeromonas sharmana* sp. nov., isolate GPTSA-6$_T$, is not a member of the genus *Aeromonas.* Int. Microbiol. 10, 61-64.

[22] Martínez-Murcia, A.J., Monera, A., Alperi, A., Figueras, M.J., Saavedra, M.J. (2009) Phylogenetic evidence suggests that strains of *Aeromonas hydrophila* subsp. *dhakensis* belong to the species *Aeromonas aquariorum* sp. nov. Curr. Microbiol. 58, 76-80.

[23] Mesbah, M., Premachandran, U., Whitman, W.B. (1989) Precise measurement of the G+C content of deoxyribonucleic acid by high-performance liquid chromatography. Int. J. Syst. Bacteriol. 39, 159-167.

[24] Miñana-Galbis, D., Farfán, M., Fusté, M.C., Lorén, J.G. (2004) *Aeromonas molluscorum* sp. nov., isolated from bivalve molluscs. Int. J. Syst. Evol. Microbiol. 54, 2073-2078.

[25] Miñana-Galbis, D., Farfán, M., Fusté, M.C., Lorén, J.G. (2007) *Aeromonas bivalvium* sp. nov., isolated from bivalve molluscs. Int. J. Syst. Evol. Microbiol. *57*, 582-587.

[26] Miñana-Galbis, D., Urbizu-Serrano, A., Farfán, M., Fusté, M.C., Lorén, J.G. (2009) Phylogenetic analisis and identification of *Aeromonas* species based on sequencing of the *cpn60* universal target. Int. J. Syst. Evol. Microbiol. 59, 1976-1983.

[27] Nhung, P.H., Hata, H., Ohkusu, K., Noda, M., Shah, M.M., Goto, K., Ezaki, T. (2007) Use of the novel phylogenetic marker *dnaJ* and DNA-DNA hybridization to clarify interrelatioships within the genus *Aeromonas*. Int. J. Syst. Evol. Microbiol. 57, 1232-1237.

[28] R Development Core Team (2008) R: A language and environment for statistical computing. [http://www.R-project.org]. ISBN 3-900051-07-0. *R Foundation for Statistical Computing,* Vienna, Austria.

[29] Saavedra, M.J., Figueras, M.J., Martínez-Murcia, A.J. (2006) Updated phylogeny of the genus *Aeromonas*. Int. J. Syst. Evol. Microbiol. 56, 2481-2487.

[30] Seshadri, R., Joseph, S.W., Chopra, A.K., Sha, J., Shaw, J., Graf, J., Haft, D., and other authors (2006) Genome sequence of *Aeromonas hydrophila* ATCC 7966^T: jack of all trades. J. Bacteriol. 188, 8272-8282.

[31] Soler, L., Yañez, M.A., Chacón, M.R., Aguilera-Arreola, M.G., Catalán, V., Figueras, M.J., Martínez-Murcia, A.J. (2004) Phylogenetic análisis of the genus *Aeromonas* based on two housekeeping genes. Int. J. Syst. Evol. Microbiol. 54, 1511-1519.

[32] Stackebrandt, E., Frederiksen, W., Garrity, G.M., Grimont, P.A.D., Kämpfer, P., Maiden, M.C.J., and other authors (2002) Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. Int. J. Syst. Evol. Microbiol. 52, 1043-1047.

[33] Sueoka, N. (1993) Directional mutation pressure, mutator mutations, and dynamics of molecular evolution. J. Mol. Evol. 37, 137-153.

[34] Xu, H.X., Kawamura, Y., Li, N., Zhao, L., Li, T.M., Li, Z.Y., Shu, S., Ezaki, T. (2000) A rapid method for determining the G+C content of bacterial chromosomes by monitoring fluorescence intensity during DNA denaturation in a capillary tube. Int. J. Syst. Evol. Microbiol. 50, 1463-1469.

FIGURE LEGENDS

Fig. 1 Tukey´s boxplot of G+C content of experimental data and of five genes used. The ends of the boxes represent the 25th and 75th percentiles. The whiskers indicate the minimum and maximum values. The horizontal bold line shows the median.

Fig. 2 Plot of experimentally determined versus the weighted average of *cpn60*, *gyrB* and *rpoD* DNA G+C content of the 31 type and reference strains of *Aeromonas* species and subspecies studied. A regression line is fitted to the data. The coefficient of determination and the regression equation are indicated.

Fig. 3 Plot of *cpn60* versus the weighted average of *cpn60*, *gyrB* and *rpoD* DNA G+C content of the 31 type and reference strains of *Aeromonas* species and subspecies studied. A regression line is fitted to the data. The coefficient of determination and the regression equation are indicated.

FIGURES

Fig. 1

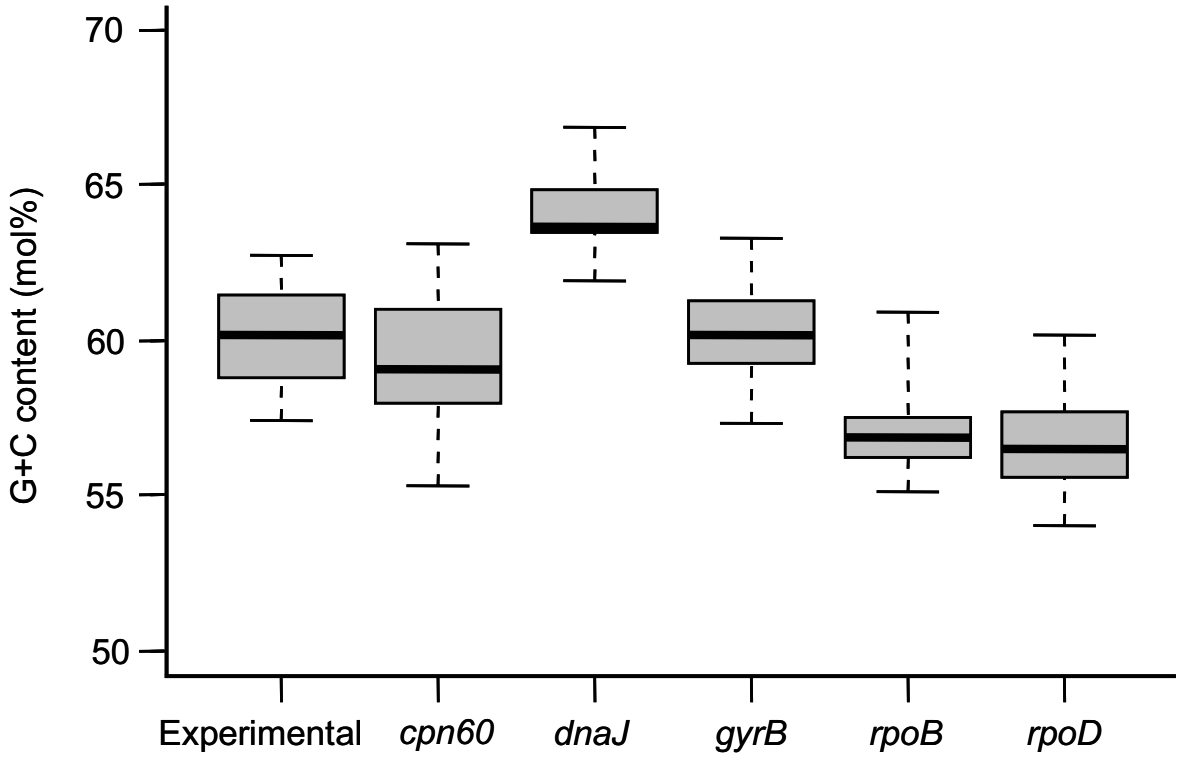


Fig. 2

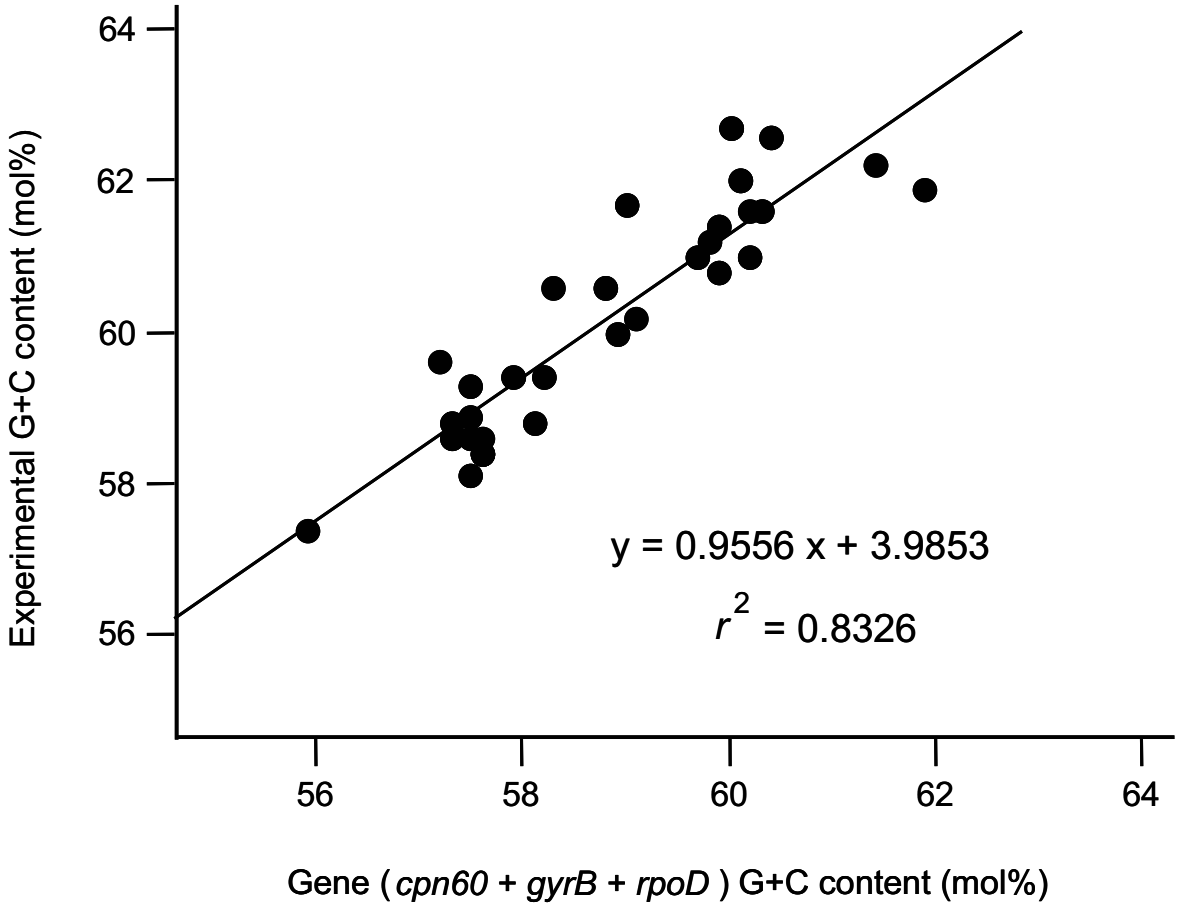
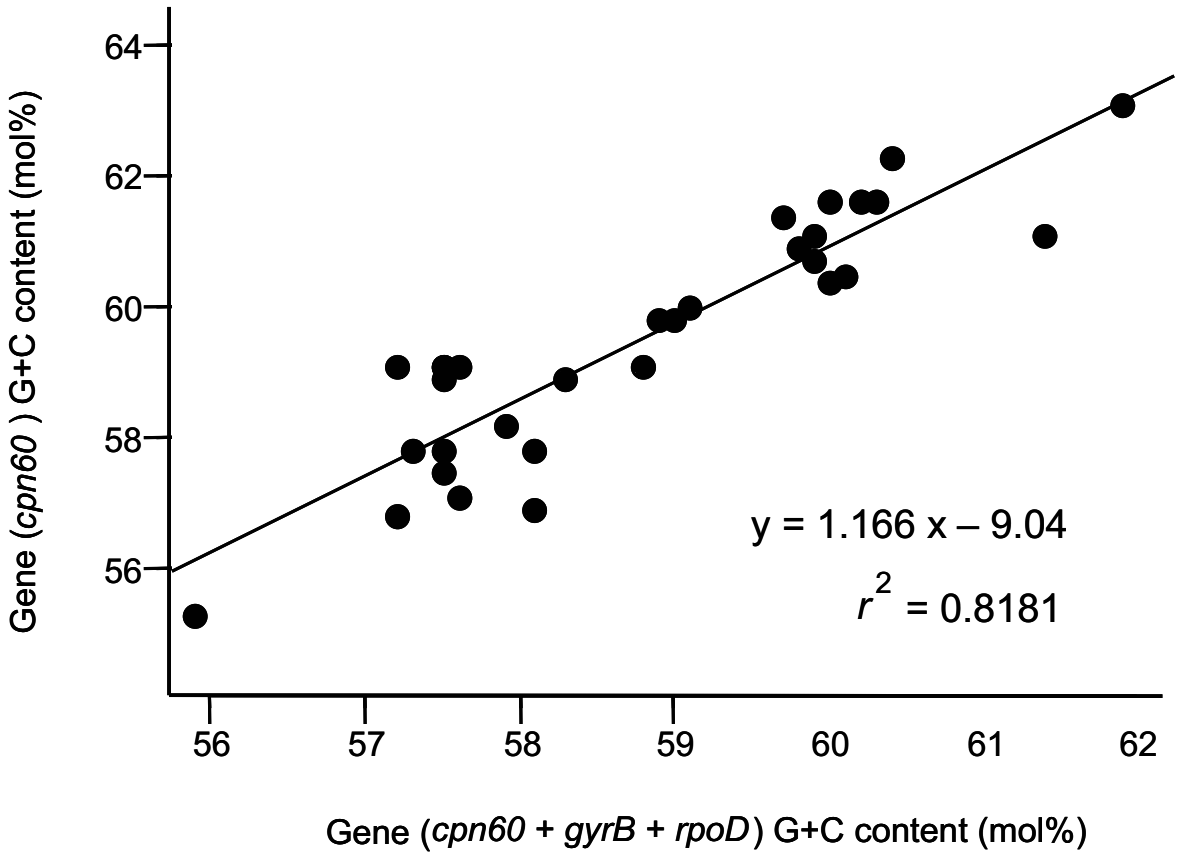

64
62
60
58
56
Experimental G+C content (mol%)
y = 0.9556 x + 3.9853
r^2 = 0.8326
56
58
60
62
64
Gene (*cpn60* + *gyrB* + *rpoD*) G+C content (mol%)

Fig. 3


Gene (*cpn60*) G+C content (mol%)
64
62
60
58
56
y = 1.166 x – 9.04
r^2 = 0.8181
56
57
58
59
60
61
62
Gene (*cpn60* + *gyrB* + *rpoD*) G+C content (mol%)

TABLES

Table 1. *Aeromonas* strains used in this study and GenBank accession numbers of gene sequences.

		GenBank accession number				
N	**Strain[a]**	***cpn60***	***dnaJ***	***gyrB***	***rpoB***	***rpoD***
1	*A. allosaccharophila* LMG 14059^T	EU306795	AB280553	AY101777	AY851132	AY169348
2	*A. aquariorum* LMG 24688^T	FJ936120	FJ936122	EU268444	FM210471	FJ936132
3	*A. bestiarum* LMG 13444^T	EU306796	AB280554	AY101774	AY851095	AY169326
4	*A. bivalvium* CECT 7113^T	EU306799	FJ936124	EF465525	EU048222	EF465512
5	*A. caviae* LMG 3775^T	EU306800	AB280555	AY101783	AY851102	AY169337
6	*A. culicicola* LMG 21852^T	EU306840	AB280556	DQ411473	AY851142	DQ411505
7	*A. encheleia* LMG 16330^T	EU306801	AB280557	AY101799	AY851133	AY169346
8	*A. enteropelogenes* LMG 12646^T	EU306837	AB280558	EF465526	EU303299	EF465508
9	*A. eucrenophila* LMG 3774^T	EU306803	AB280559	AY101776	AY851116	AY169339
10	*A. hydrophila* subsp. *dhakensis* LMG 19562^R	EU306806	AB280560	AM262163	DQ448289	EF465510
11	*A. hydrophila* subsp. *hydrophila* LMG 2844^T	EU306804	AB280561	AY101778	AY851091	AY169325
12	*A. hydrophila* subsp. r*anae* LMG 19707^R	EU306805	AB280562	AM262162	DQ448290	EF465509
13	*A. ichthiosmia* LMG 12645^T	EU306841	AB280563	EF465527	EU313542	AY169342
14	*A. jandaei* LMG 12221^T	EU306807	AB280564	AY101780	AY851121	AY169341
15	*A. media* LMG 9073^T	EU306808	AB280565	AY101782	AY851112	AY169338
16	*A. molluscorum* CECT 5864^T	EU306811	AB280566	EF465521	DQ448280	EF465515
17	*A. popoffii* LMG 17541^T	EU306814	AB280567	AY101801	AY851138	AY169347
18	*A. salmonicida* subsp. *achromogenes* LMG 14900^R	EU306824	AB280568	AY101785	DQ448285	AY169329
19	*A. salmonicida* subsp. *masoucida* LMG 3782^R	EU306825	AB280569	AY101784	DQ448287	AY169330
20	*A. salmonicida* subsp. *pectinolytica* LMG 19569^R	EU306827	AB280570	AY101810	DQ448288	AY169324
21	*A. salmonicida* subsp. s*almonicida* LMG 3780^T	EU306828	AB280571	AY101773	AY851098	AY169327
22	*A. salmonicida* subsp. *smithia* LMG 20223^R	EU306829	AB280572	AM262159	DQ448286	AY169331
23	*A. schubertii* LMG 9074^T	EU306830	AB280574	AY101772	AY851129	AY169336
24	*A. simiae* LMG 22269^T	EU306833	AB280573	DQ411480	AY851143	DQ411508
25	*A. sobria* LMG 3783^T	EU306834	AB280575	AY101781	AY851119	AY169340
26	*A. tecta* DSM 17300^T	FJ936121	FJ936130	AJ964952	FJ936131	FJ936133
27	*A. trota* LMG 12223^T	EU306836	AB280576	AY101800	AY851131	AY169344
28	*A. veronii* bv. Sobria LMG 3785^R	EU306838	AB280578	AY101775	AY851120	AY169333
29	*A. veronii* bv. Veronii LMG 9075^T	EU306839	AB280577	AY101795	AY851122	AY127862
30	*Aeromonas* sp. HG11 LMG 13075^R	EU306802	AB280552	AY101779	AY851127	AY169343
31	*Aeromonas* sp. HG13 LMG 17321^R	EU306835	FJ936129	AY101806	AY851130	AY169345

[a] CECT, Spanish type culture collection; DSM, German collection of microorganisms and cell cultures; LMG, Belgian co-ordinated collections of microorganisms; ^{T}, type strain; ^{R}, reference strain; HG, DNA hybridization group

Table 2. Comparison of the mol% G+C content within *Aeromonas* genus obtained from HPLC method, calculated based on housekeeping genes method, and previously reported in literature.

	mol% G+C content					
***Aeromonas* strains**	**Experimental (HPLC)**	**Predicted *cpn60, gyrB, rpoD*[a]**	**Dif[b]**	**Predicted *cpn60*[a]**	**Dif[b]**	**Published[c]**
A. allosaccharophila	58.9	59.0 ± 0.15	0.17	59.0 ± 0.21	0.14	59.5 [1]
A. aquariorum	61.0	61.3 ± 0.15	0.49	60.8 ± 0.17	0.22	
A. bestiarum	60.6	60.2 ± 0.11	0.66	59.9 ± 0.16	0.69	
A. bivalvium	62.6	61.7 ± 0.17	1.44	62.1 ± 0.29	0.55	62.6 [2]
A. caviae	61.6	61.6 ± 0.16	0.00	61.6 ± 0.24	0.02	61 - 63 [1]
A. culicicola	58.8	58.8 ± 0.16	0.00	59.0 ± 0.21	0.24	
A. encheleia	62.7	61.3 ± 0.15	2.23	61.6 ± 0.24	1.12	59.4 - 60.8 [1]
A. enteropelogenes	60.0	60.3 ± 0.11	0.50	60.4 ± 0.16	0.38	
A. eucrenophila	61.0	61.1 ± 0.13	0.16	61.5 ± 0.23	0.45	59.8 - 62.6 [1]
A. hydrophila subsp. *dhakensis*	62.0	61.4 ± 0.15	0.97	60.9 ± 0.18	1.15	
A. hydrophila subsp. *hydrophila*	61.4	61.2 ± 0.14	0.33	61.0 ± 0.19	0.42	58 - 62 [1]; 61.5 [3]
A. hydrophila subsp. *ranae*	61.7	60.4 ± 0.11	2.11	60.4 ± 0.16	1.32	
A. ichthiosmia	59.3	59.0 ± 0.15	0.51	58.8 ± 0.23	0.46	
A. jandaei	58.8	59.5 ± 0.12	1.19	59.0 ± 0.21	0.24	
A. media	60.8	61.2 ± 0.14	0.66	61.3 ± 0.21	0.45	62.3 [1]
A. molluscorum	59.4	59.3 ± 0.13	0.17	59.3 ± 0.19	0.09	59.0 – 59.4 [4]
A. popoffii	59.4	59.5 ± 0.12	0.17	58.4 ± 0.27	0.96	57.7 - 59.6 [1]
A. salmonicida subsp. *achromogenes*	58.6	59.0 ± 0.15	0.68	59.9 ± 0.16	1.31	57- 59 [1]
A. salmonicida subsp. *masoucida*	58.1	59.0 ± 0.15	1.55	59.9 ± 0.16	1.81	
A. salmonicida subsp. *pectinolytica*	58.4	59.0 ± 0.15	1.03	59.8 ± 0.16	1.38	
A. salmonicida subsp. *salmonicida*	58.4	59.0 ± 0.15	1.03	59.9 ± 0.16	1.51	57- 59 [1]
A. salmonicida subsp. *smithia*	58.6	58.7 ± 0.17	0.17	59.9 ± 0.16	1.31	55.9 [1]
A. schubertii	61.9	63.2 ± 0.27	2.10	62.6 ± 0.35	0.68	
A. simiae	61.2	61.2 ± 0.14	0.00	61.1 ± 0.20	0.09	
A. sobria	57.4	57.4 ± 0.25	0.00	57.4 ± 0.39	0.03	58 - 60 [1]
A. tecta	60.2	60.5 ± 0.11	0.50	60.5 ± 0.16	0.31	
A. trota	60.6	59.7 ± 0.12	1.49	59.8 ± 0.16	0.82	
A. veronii bv. Sobria	58.6	59.0 ± 0.15	0.68	58.6 ± 0.26	0.03	
A. veronii bv. Veronii	59.6	58.7 ± 0.17	1.51	58.4 ± 0.28	1.23	57.6 - 58.2 [1]
Aeromonas sp. HG11	61.6	61.5 ± 0.16	0.16	61.6 ± 0.24	0.02	
Aeromonas sp. HG13	62.2	62.7 ± 0.24	0.80	61.3 ± 0.21	0.95	

[a] Predicted G+C content ± standard deviation

[b] Absolute differences between experimental and calculated G+C content

[c] Sources: [1][19]; [2][25]; [3][30]; [4][24]

Table 3. Regression parameters comparison for the estimation of genomic DNA G+C content.

Genes[a]	nt[b]	Regression equation	*r*[c]	*P*[d]	AIC[e]
cpn60	555.0	y = 0.6687 x + 20.3882	0.8228	$1.348\ 10^{-8}$	-7.044
dnaJ	849.9	y = 0.9455 x – 0.4465	0.7066	$8.865\ 10^{-6}$	6.536
gyrB	1001.0	y = 0.9837 x + 0.9584	0.8601	$5.620\ 10^{-10}$	-13.751
rpoB	516.7	y = 0.7291 x + 19.0900	0.6102	$2.671\ 10^{-4}$	13.542
rpoD	820.5	y = 0.7927 x + 15.3058	0.8529	$1.114\ 10^{-9}$	-12.305
wm3	-	y = 0.9560 x + 3.9819	0.9124	$8.906\ 10^{-13}$	-27.419

[a] wm3, weighted mean of three genes (*cpn60*, *gyrB* and *rpoD*)

[b] nt, mean number of nucleotides. In all cases, except to *cpn60*, the lenght of the sequences analyzed was distinct for the different species or subspecies.

[c] *r*, Pearson´s product-moment correlation coefficient

[d] Statistical significance

[e] AIC, Akaike´s Information Criterion

Table 4. Predictions of genomic DNA G+C content from the three genes (*cpn60*, *gyrB* and *rpoD*) or using *cpn60* gene.

		***cpn60* + *gyrB* + *rpoD*[a]**		***cpn60***	
Strains	**Experimental**	**Calculated**	**Dif[c]**	**Calculated**	**Dif[c]**
A. molluscorum 93M	59.4	59.5	0.1	59.2	0.2
A. molluscorum 431T	59.0	59.4	0.4	59.1	0.1
A. molluscorum 849T	59.3	59.9	0.6	59.5	0.2
A. molluscorum 869N	59.3	59.4	0.1	59.1	0.2
	Genomic[b]				
A. hydrophila ATCC 7966^T	61.5	60.9	0.6	61.0	0.5
A. salmonicida A449	58.5	58.5	0.0	59.9	1.4

[a] GenBank accession numbers of the nucleotide sequences used of each strain of *A. molluscorum*: EU306809 (*cpn60*, strain 93M); EU306810 (*cpn60*, strain 431T); EU306812 (*cpn60*, strain 849T); EU306813 (*cpn60*, strain 869N); EF465519 (*gyrB*, strain 93M); EF465520 (*gyrB*, strain 431T); EF465522 (*gyrB*, strain 849T); EF465523 (*gyrB*, strain 869N); EF465513 (*rpoD*, strain 93M); EF465514 (*rpoD*, strain 431T); EF465516 (*rpoD*, strain 849T); EF465517 (*rpoD*, strain 869N).

[b] Data obtained from the whole genomes of *A. hydrophila* ATCC 7966^T and *A. salmonicida* A449 (GenBank accession numbers: CP000462 and CP000644, respectively).

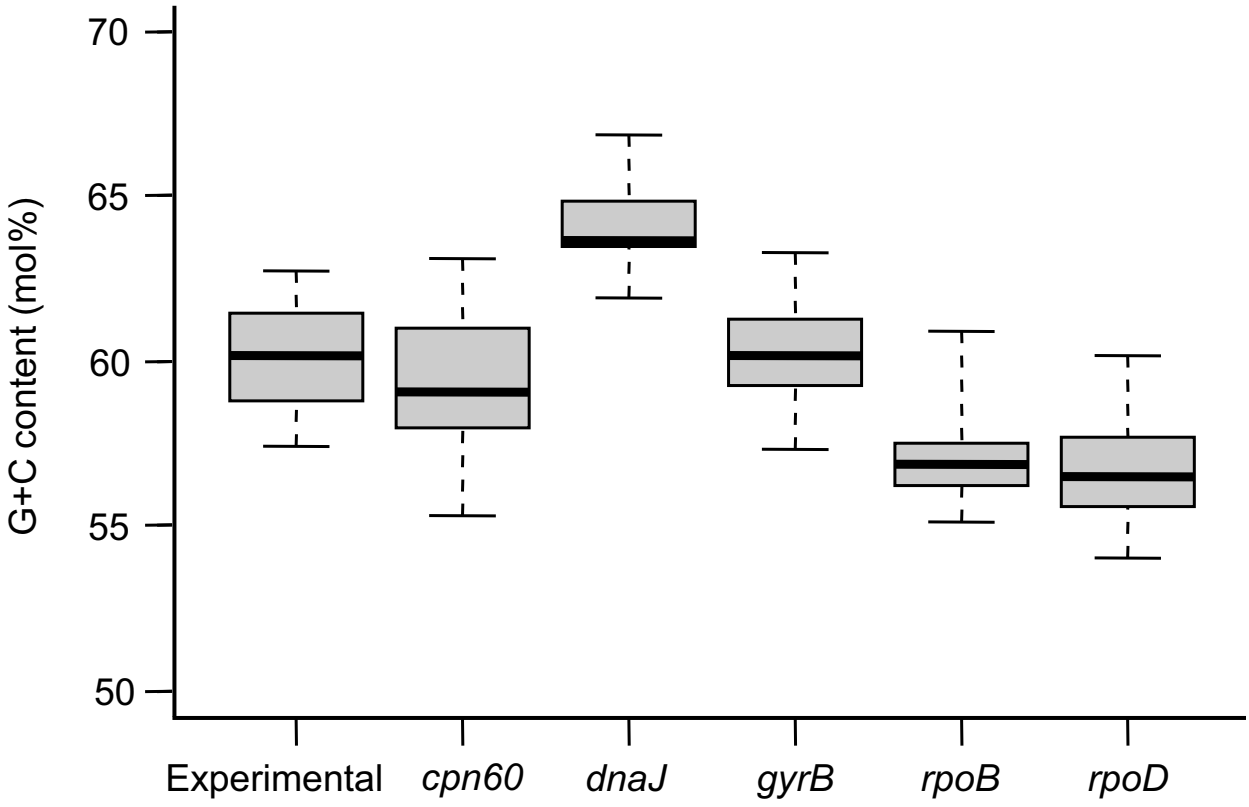
[c] Absolute differences between experimental or genomic and calculated G+C content

Table 5. Intraspecific variation of the G+C content calculated from gene sequences within *Aeromonas* species.

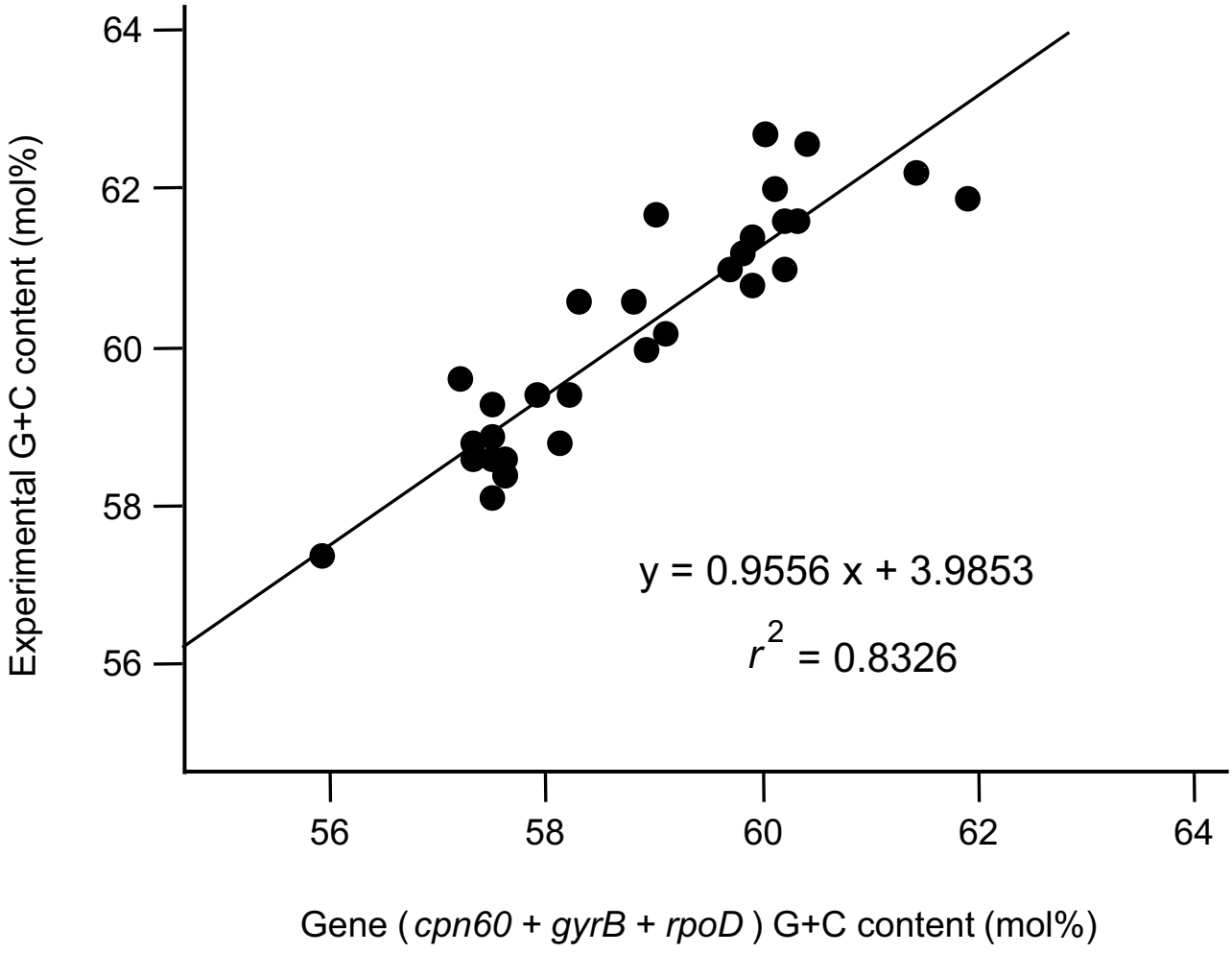
	Mean G+C content[a]		
***Aeromonas* species**	***cpn60***	***gyrB***	***rpoD***
A. bestiarum	59.6 ± 0.1 (13)	60.6 ± 0.1 (7)	56.0 ± 0.2 (7)
A. hydrophila	60.2 ± 0.2 (8)	60.7 ± 0.1 (10)	57.6 ± 0.1 (6)
A. molluscorum	59.2 ± 0.4 (5)	59.3 ± 0.1 (5)	55.7 ± 0.1 (5)
A. salmonicida	59.2 ± 0.1 (13)	59.4 ± 0.1 (8)	54.1 ± 0.1 (8)

[a] Mean G+C content (mol%) ± standard error. The number of strains used is given in parentheses

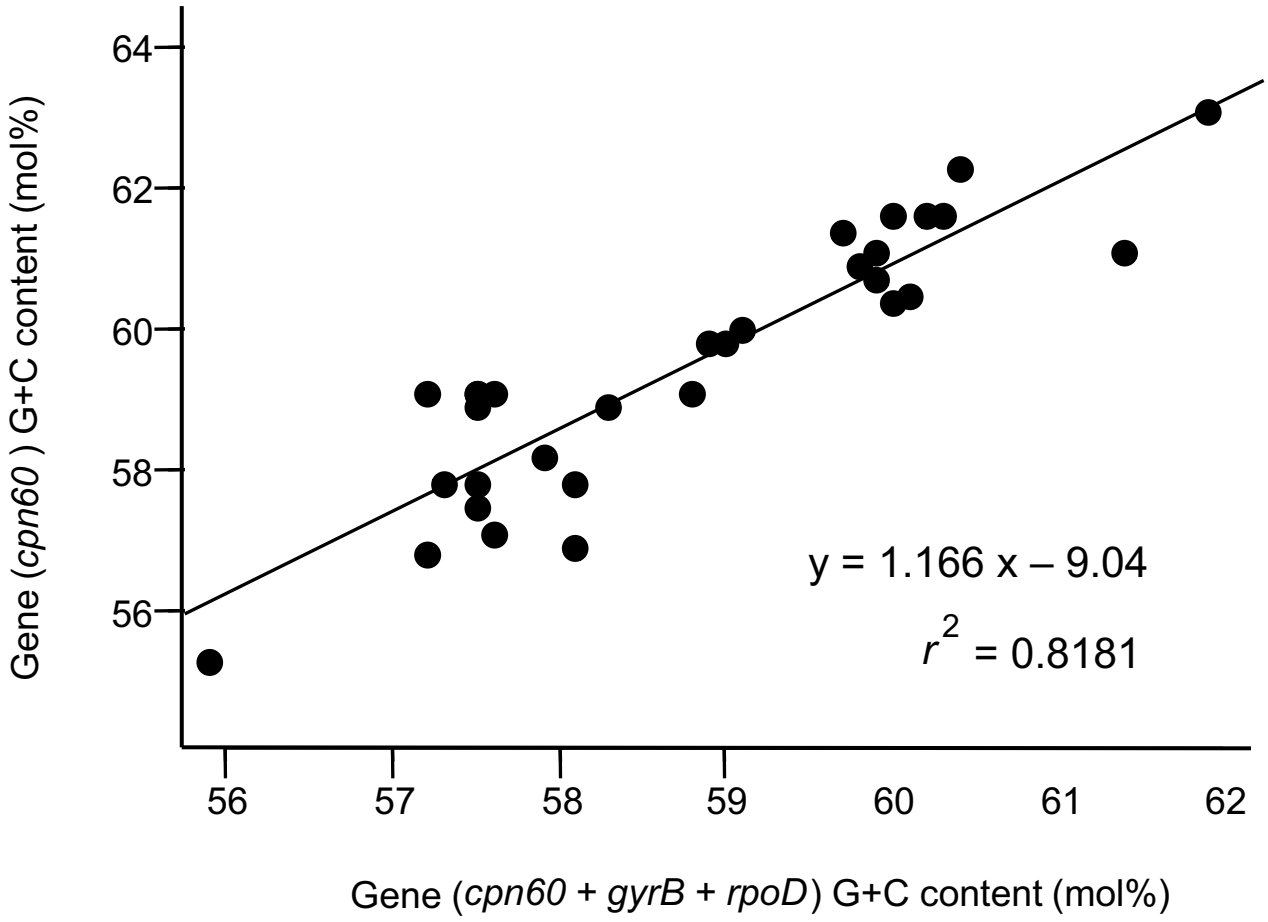
Figure


70
65
60
55
50
G+C content (mol%)
Experimental
cpn60
dnaJ
gyrB
rpoB
rpoD

Figure


64
62
60
58
56
Experimental G+C content (mol%)
y = 0.9556 x + 3.9853
r 2 = 0.8326
56
58
60
62
64
Gene (cpn60 + gyrB + rpoD) G+C content (mol%)

Figure


64
62
60
58
56
Gene (cpn60) G+C content (mol%)
y = 1.166 x – 9.04
r2 = 0.8181
56
57
58
59
60
61
62
Gene (cpn60 + gyrB + rpoD) G+C content (mol%)