Frequency of quorum sensing mutations in *Pseudomonas aeruginosa* strains isolated from different environments

Kathleen O'Connor, Conan Y. Zhao & Stephen P. Diggle*

Center for Microbial Dynamics & Infection, School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA 30332.

*Correspondence: stephen.diggle@biosci.gatech.edu

Keywords: *Pseudomonas aeruginosa*, quorum sensing, LasR, cystic fibrosis

Abstract

Pseudomonas aeruginosa uses quorum sensing (QS) to coordinate the expression of multiple genes necessary for establishing and maintaining infection. *lasR* QS mutations have been shown to frequently arise in cystic fibrosis (CF) lung infections, however, there has been far less emphasis on determining whether QS system mutations arise across other environments. To test this, we utilized 852 publicly available sequenced *P. aeruginosa* genomes from the *Pseudomonas* International Consortium Database (IPCD) to study *P. aeruginosa* QS mutational signatures. We found that across all isolates, LasR is the most variable protein sequence compared to other QS proteins. In order to study isolates by source, we focused on a subset of 654 isolates collected from CF, wounds, and non-infection environmental isolates, where we could clearly identify their source. Using this sub-set analysis, we found that LasR mutations are not specific to CF lungs, but are common across all environments. We then used amino acid length as a proxy for observing loss of function in LasR proteins among the strains. We found that truncated LasR proteins are more abundant in *P. aeruginosa* strains isolated from human infection than the environment. Overall, our findings suggest that the evolution of *lasR* QS mutations in *P. aeruginosa* are common and not limited to infection environments.

Introduction

Pseudomonas aeruginosa is a Gram-negative opportunistic pathogen equipped with a large genome, which enables it to be metabolically versatile and capable of occupying a range of different habitats, especially human and animal impacted environments [1, 2]. It is intrinsically resistant to many classes of antibiotic, and it produces a range of tissue damaging extracellular products such as exoenzymes and phenazine pigments in order to aid its dissemination and spread within a host [2]. *P. aeruginosa* is one of the most prominent bacterial pathogens that colonizes cystic fibrosis (CF) lungs and it is often the dominant CF lung pathogen, particularly as the infection becomes more chronic over time [3]. In addition, *P. aeruginosa* frequently infects chronic wounds, often in conjunction with other microbial species [4].

One of the major adaptations of *P. aeruginosa* during chronic infection is the loss of quorum sensing (QS). QS in *P. aeruginosa* regulates the expression of hundreds of genes, including those that encode for secreted products and virulence factors [5, 6]. In *P. aeruginosa,* QS is regulated by a complex hierarchical network of genes, composed of two complete *N*-acyl homoserine lactone (AHL) circuits, LasR-LasI and RhlR-RhlI, as well as an orphan regulator termed QscR [5, 7]. The Las and Rhl systems are composed of LuxR-LuxI pairs, homologous to other bacterial QS systems. The LuxR-type receptors (LasR, RhlR) act as transcriptional regulators, and the LuxI-type proteins (LasI, RhlI) are signal synthases. LasI produces 3-oxo-dodecanoyl-L-homoserine lactone (3OC12-HSL), and RhlI produces butanoyl-L-homoserine lactone (C4-HSL) [5]. Both signals can function in a combinatorial manner to regulate certain genes [8, 9]. Working in conjunction with the two AHL systems is an alkyl-quinolone (AQ) system, comprising the *pqsABCDE* operon and *pqsH, pqsL* and *pqsR* (*mvfR*) genes. These genes drive the synthesis and response of 2-heptyl-3-hydroxy-4-quinolone (the Pseudomonas quinolone signal PQS), which is used as a QS signal and which also has iron chelating properties [10, 11].

LasR was first identified in 1991 as a regulator of the *lasB* (elastase) gene [12]. It has since been described as a key QS regulator in the well-studied laboratory strains PAO1 and PA14 [2], where it has been shown to sit at the top of the QS hierarchy, regulating both the *rhl* and *pqs* systems [5, 6]. *lasR* mutants have frequently been isolated from CF lungs [13-16] and more recently, some CF strains use RhlR to regulate the *rhl* and *pqs* systems in the absence of a functional LasR [6, 16-19]. The decoupling of the AHL QS hierarchy requires the inactivation of MexT, a regulator of the multi-drug efflux pump operon MexEF-OprN [18, 20]. PqsE and RhlR have also been shown to function as a ligand:receptor pair [21].

The ecological and evolutionary implications of QS re-wiring remain to be explored, and the drivers for *lasR* mutation before and during infection are unknown. To date, little is known about QS mutations outside of an infection environment, and so in this study, we explored the diversity and frequency of QS mutations across a range of ecologically distinct environments in order to determine (i) which QS genes are frequently mutated; (ii) mutational signatures, or patterns in QS gene mutation specific to isolate source.

Results

We utilized the published sequences of 852 *P. aeruginosa* isolates from the International *Pseudomonas* Consortium Database (IPCD); a database representing a range of *P. aeruginosa* strains from different sources including rivers, infections and plants [22]. We queried a number of key QS genes from the *las*, *rhl and pqs* systems against gene sequences from PAO1 using BLASTn for all 852 isolates. We determined the putative amino acid sequence for each gene and calculated dissimilarity scores using BLOSUM80, an empirical amino acid substitution matrix [13]. All analyses were conducted in R version 4.3.

We first looked at the number of sequences we had for each QS protein, and the diversity of the protein sequences. When we queried each QS gene nucleotide sequence against the 852 isolates, the query returned less than 852 sequences for each gene. This disparity is likely due to gaps in sequences, gene deletions, and extensive mutations, preventing BLASTn from returning a query. Fig. 1A shows that a LasR query returned the fewest number of sequences, suggesting that there are many strains that contain large deletions in LasR, truncations, or are missing the LasR gene entirely. After translating the sequences, we found that LasR also had the most unique protein sequences across 852 isolates (Fig. 1B). Given this finding, we analyzed whether there was a mutational signature for the *las* system in order to determine whether certain kinds of mutation or divergence were specific to *las* genes, and if these mutations were specific to isolate source. We created PCA plots of LasR (Fig. 1C) and LasI (Fig. 1D) proteins from all returned isolates and found that the LasR protein was distributed across the PCA plot, and the most divergent strains for LasR were truncated. Compared to LasR, the other key QS proteins were more conserved across isolates.

After analyzing all QS proteins, we then specifically focused on LasR. To determine if there were LasR mutations specific to each environment, we categorized the strains by source. Using data from the IPCD, we selected a subset of 654 strains labeled as “environmental”, “cystic fibrosis” or “CF”, and “wound” or “ulcer” or “burn” and reclassified them as environmental (209 strains), CF (396 strains), or wound (wound, ulcer and burn) (49 strains); 654 total. The remaining 198 strains from the original set of 852 strains were of uncertain origin and therefore not used in this particular analysis. To establish a threshold by which a protein could be deemed functional or not, we looked at truncated LasR proteins within each environment. We compared the amino acid length of LasR in the IPCD strains to the PAO1 LasR protein - which is equal in length to many commonly researched strains including PA14, PAK and an epidemic CF strain, LESB58. Our assumption was that a truncated protein due to shortened DNA sequence or an early stop site, would lead to a nonfunctional protein. We used a stringent 100% length as a cut-off, and any protein shorter than full-length was considered truncated. Fig. 2A shows the proportion of each group that had truncated LasR proteins with CF, environmental, and wound isolates having 20%, 11% and 30% truncations respectively. We also used a PCA plot to visualize LasR amino acid variation by environment using BLOSUM80 generated dissimilarity scores (Fig. 2B). Overall, we found that *lasR* mutations are ubiquitous across all environments, but there is a larger percentage of strains with truncated LasR proteins found in infection environments.

Discussion

In *P. aeruginosa*, *lasR* QS mutants are frequently isolated from human chronic infection, but it has remained unclear whether such mutants specifically evolve in infection environments or are common across multiple environments. Using a publicly available database of 852 fully sequenced isolates from CF, wounds and non-infection based (environmental) isolates, we determined the frequency and pattern of *lasR* and other QS mutations in *P. aeruginosa*. We found that (i) LasR is the most variable protein of all the major QS proteins; (ii) *lasR* mutations are found in isolates across all environments, suggesting that any environment can drive the evolution of these mutations.

But what does drive the evolution of *lasR* mutations and what fitness benefits do *lasR* mutations provide to *P. aeruginosa* isolates or populations? First, *lasR* mutants could arise in populations through social cheating, where mutants exploit the social interactions and exoproducts produced by *lasR* intact cells [23, 24]. Controlled experiments have shown that *lasR* mutants can socially exploit wild type cells *in vitro* [24] and *in vivo* [25], although it is unclear whether the spatial structuring found within infections will allow the close proximity of different isolates to allow for regular cheating. Importantly, QS genes have recently been shown to be down-regulated during infection compared to *in vitro* conditions*,* questioning the long-held belief that a functional QS system is essential for *P. aeruginosa* to establish and persist in human infections [6, 26]. This would likely reduce any fitness benefits of being a *lasR* mutant persisting via social cheating. Second, *lasR* mutants may have increased fitness in particular environments due to certain phenotypes driven by the mutation being beneficial. For example, *lasR* mutations have previously been shown to confer a growth advantage with particular carbon and nitrogen sources, including amino acids [27]. Third, *lasR* mutants may be more competitive than *lasR* positive cells, which provides fitness benefits against other *P. aeruginosa* strains or other species.

There are, however, likely evolutionary benefits for both the maintenance and loss of LasR so that both *lasR* positive and negative strains can stably coexist in heterogenous populations and contribute to an overall community function. In recent support of this idea, it has been shown that (i) *lasR*- strains overproduce Rhl-associated factors and cross-feed wild type cells in low iron environments, which will likely impact infection dynamics of mixed populations [28]; (ii) mixed *lasR* +/- populations display decreased virulence in mouse models of infection [25]; (iii) mixed populations exhibit enhanced tolerance to beta-lactam antibiotics [29]. Taken together, this suggests there are likely to be considerable fitness advantages to cells growing in heterogeneous QS populations, perhaps as a bet-hedging mechanism for future disturbance events.

Overall, our work highlights that *lasR* mutations are the most commonly found QS mutation across different environments, although we do not know whether mutations in the *lasR* gene always results in a loss of QS function. Indeed, recent studies on QS in *P. aeruginosa* has revealed that the complex and intertwined *las*, *rhl*, *and pqs* systems can be rewired in the event that *lasR* becomes mutated [16, 18][9, 19]. It is not always clear whether these strains are entirely QS-null or if they have re-wired their QS systems to circumvent the loss of *lasR*. Further work is needed to determine why strains lose functional LasR proteins, and what fitness benefits the strains or community gains. Future work should more strongly focus on the ecology of mixed QS-phenotypes to better understand QS-involvement in infection and other environments. With ongoing work identifying QS-inhibitors targeting the *las* QS system, the frequency of *lasR* mutated strains found in our study suggests that this particular pursuit is likely to fail.

Materials and Methods

Querying QS genes from the International Pseudomonas Consortium Database. Using nucleotide sequences from PAO1, we queried QS genes (see Fig. 1) using BLASTn for isolates from the IPCD [12]. We chose this strain because it is a fully sequenced, frequently used lab strain. We then translated these sequences into protein sequences calculating putative amino acid sequence similarities using BLOSUM80 [13]. First, we compared genes found in each isolate against our reference strain, PAO1, normalized against the similarity of the reference against itself. We then calculated the mean dissimilarity score of all isolates compared to PAO1. Some isolates were missing genes due to sequencing errors or true truncations, the number of isolates with a given gene present was under 852 for all genes. All analyses, including translation steps were conducted in R version 4.3. All code and files are available on Github (https://github.gatech.edu/login?return_to=https%3A%2F%2Fgithub.gatech.edu%2Fkoconnor36%2FFrequency_of_quorum_sensing_mutations_in_Pa2021).

Creating an IPCD database using BLASTn. We pulled IPCD data from GenBank. We used the makeblastdb/ command to generate a database of all isolate contigs.

Using BLASTn to find QS genes for each isolate. Using our generated database, we queried the PAO1 sequence from each gene, found from Genbank, against the database. We generated csv files for each gene which included the gene sequences for each isolate.

Translating nucleotide to amino acid sequence. We translated genes to proteins using a custom R script. We first queried only for sequences starting with a canonical ATG start codon. We exclude sequences with fully unresolvable nucleotides (coded as “-“), but allowed fuzzy codons so long as they resolved to unambiguous amino acids. We translated the sequences meeting these criteria using the translate function from the BioStrings R package (v.2.58.0).

Calculating dissimilarity scores for isolates’ QS proteins. All sequence analyses were performed in R (v.4.0.2) using the Biostrings package v.2.58.0. We compared isolate protein sequences to PAO1 protein sequences using BLOSUM80, a matrix designed to compared protein sequences within species. We found that close to 50% of all isolate proteins were identical to PAO1.

Determining truncation rates for LasR and categorizing isolates by location. We determined the length of the reference LasR protein, from PAO1, compared to each isolate protein. If the isolate protein was 100% or less of the length of the PAO1 protein, we categorized it as truncated. Sequences were categorized as CF-originated (CF), environmental (ENV), or wound (WND). If the sequence was entered into IPCD as environmental, we adopted that label. Additionally, we included sequences labeled from animal hosts as environmental. For CF, we only included sequences with sources explicitly labeled as CF. For wound, we included sequences labeled as wound, ulcer, and burn.

Author contributions. SPD and KOC designed the study. KOC and CYZ performed the *in silico* analysis of the data. All authors contributed to the writing of the manuscript.

Competing interests. The authors declare no competing interests.

Funding and acknowledgements. We wish to thank The Cystic Fibrosis Foundation (DIGGLE18I0 and DIGGLE20G0) to SPD for funding. We also thank members of the Diggle Lab and Jon Gerhart for helpful discussion.

Figure legends

Figure 1. Determining variability in QS proteins between *P. aeruginosa* isolates from the IPCD. (A) We created a database of 852 isolates and used PAO1 to search for the QS proteins of each isolate. Due to the variation in each isolate's genome and due to gaps in sequencing, each protein queried returned fewer than 852 sequences (shown in gray). We also determined the number of unique sequences for each protein and found that LasR had the highest number of unique sequences (in color). (B) Using a custom dissimilarity metric (BLOSUM80), we calculated mean dissimilarity scores. We found that LasR had the highest mean dissimilarity score compared to all QS proteins, and the largest variation. We conducted principal component analyses (PCA) and plotted the similarity scores of all isolates for the LasR protein (C) and the LasI protein (D).

Figure 2. LasR truncations are found in isolates from all sources. To observe the fraction of truncated proteins across all environments, we categorized the isolates into 3 groups: cystic fibrosis (CF), environmental (ENV) or wound (WND). (A) We show the number of truncated proteins out of the total number of isolates in each group. The PCA plot (B) depicts the similarity scores for all groups compared, and we see the arm of truncated proteins consisted primarily of CF and WND isolates.

References

1. **Rahme LG, Stevens EJ, Wolfort SF, Shao J, Tompkins RG et al.** Common virulence factors for bacterial pathogenicity in plants and animals. *Science* 1995;268(5219):1899-1902.
2. **Diggle SP, Whiteley M**. Microbe Profile: *Pseudomonas aeruginosa*: opportunistic pathogen and lab rat. *Microbiology* 2020;166(1):30-33.
3. **Elborn JS**. Cystic fibrosis. *The Lancet* 2016;388(10059):2519-2531.
4. **Ibberson CB, Whiteley M**. The social life of microbes in chronic infection. *Curr Opin Microbiol* 2020;53:44-50.
5. **Whiteley M, Diggle SP, Greenberg EP**. Progress in and promise of bacterial quorum sensing research. *Nature* 2017;551(7680):313-320.
6. **Azimi S, Klementiev AD, Whiteley M, Diggle SP.** Bacterial Quorum Sensing During Infection. *Annu Rev Microbiol* 2020;74:201-219.
7. **Schuster M, Sexton DJ, Diggle SP, Greenberg EP**. Acyl-homoserine lactone quorum sensing: from evolution to application. *Annu Rev Microbiol*, Review 2013;67(1):43-63.
8. **Cornforth DM, Popat R, McNally L, Gurney J, Scott-Phillips TC et al.** Combinatorial quorum sensing allows bacteria to resolve their social and physical environment. *Proc Natl Acad Sci U S A* 2014;111(11):4280-4284.
9. **Gurney J, Azimi S, Brown SP, Diggle SP**. Combinatorial quorum sensing in Pseudomonas aeruginosa allows for novel cheating strategies. *Microbiology (Reading)* 2020;166(8):777-784.
10. **Dubern JF, Diggle SP**. Quorum sensing by 2-alkyl-4-quinolones in *Pseudomonas aeruginosa* and other bacterial species. *Mol Biosyst*, Review 2008;4(9):882-888.
11. **Diggle SP, Matthijs S, Wright VJ, Fletcher MP, Chhabra SR et al.** The *Pseudomonas aeruginosa* 4-quinolone signal molecules HHQ and PQS play multifunctional roles in quorum sensing and iron entrapment. *Chem Biol* 2007;14(1):87-96.
12. **Gambello MJ, Iglewski BH**. Cloning and characterization of the *Pseudomonas aeruginosa lasR* gene, a transcriptional activator of elastase expression. *J Bacteriol* 1991;173(9):3000-3009.
13. **Smith EE, Buckley DG, Wu Z, Saenphimmachak C, Hoffman LR et al.** Genetic adaptation by *Pseudomonas aeruginosa* to the airways of cystic fibrosis patients. *Proc Natl Acad Sci U S A* 2006;103(22):8487-8492.
14. **Ciofu O, Mandsberg LF, Bjarnsholt T, Wassermann T, Hoiby N**. Genetic adaptation of *Pseudomonas aeruginosa* during chronic lung infection of patients with cystic fibrosis: strong and weak mutators with heterogeneous genetic backgrounds emerge in *mucA* and/or *lasR* mutants. *Microbiology* 2010;156(Pt 4):1108-1119.
15. **Jiricny N, Molin S, Foster K, Diggle SP, Scanlan PD et al.** Loss of social behaviours in populations of *Pseudomonas aeruginosa* infecting lungs of patients with cystic fibrosis. *PLoS One* 2014;9(1):e83124.
16. **Feltner JB, Wolter DJ, Pope CE, Groleau MC, Smalley NE et al.** LasR Variant Cystic Fibrosis Isolates Reveal an Adaptable Quorum-Sensing Hierarchy in *Pseudomonas aeruginosa*. *mBio* 2016;7(5):e01513-01516-01519.
17. **Diggle SP, Winzer K, Chhabra SR, Worrall KE, Camara M et al.** The *Pseudomonas aeruginosa* quinolone signal molecule overcomes the cell density-dependency of the quorum sensing hierarchy, regulates *rhl*-dependent genes at the onset of stationary phase and can be produced in the absence of LasR. *Mol Microbiol* 2003;50(1):29-43.

18. **Kostylev M, Kim DY, Smalley NE, Salukhe I, Greenberg EP et al.** Evolution of the *Pseudomonas aeruginosa* quorum-sensing hierarchy. *Proc Natl Acad Sci U S A* 2019;116(14):7027-7032.

19. **Cruz RL, Asfahl KL, Van den Bossche S, Coenye T, Crabbe A et al.** RhlR-Regulated Acyl-Homoserine Lactone Quorum Sensing in a Cystic Fibrosis Isolate of *Pseudomonas aeruginosa*. *mBio* 2020;11(2).

20. **Oshri RD, Zrihen KS, Shner I, Omer Bendori S, Eldar A**. Selection for increased quorum-sensing cooperation in *Pseudomonas aeruginosa* through the shut-down of a drug resistance pump. *ISME J* 2018;12(10):2458-2469.

21. **Mukherjee S, Moustafa DA, Stergioula V, Smith CD, Goldberg JB et al.** The PqsE and RhlR proteins are an autoinducer synthase-receptor pair that control virulence and biofilm development in *Pseudomonas aeruginosa*. *Proc Natl Acad Sci U S A* 2018;115(40):E9411-E9418.

22. **Freschi L, Vincent AT, Jeukens J, Emond-Rheault JG, Kukavica-Ibrulj I et al.** The *Pseudomonas aeruginosa* Pan-Genome Provides New Insights on Its Population Structure, Horizontal Gene Transfer, and Pathogenicity. *Genome Biol Evol* 2019;11(1):109-120.

23. **West SA, Griffin AS, Gardner A, Diggle SP**. Social evolution theory for microorganisms. *Nat Rev Microbiol*, Review 2006;4(8):597-607.

24. **Diggle SP, Griffin AS, Campbell GS, West SA**. Cooperation and conflict in quorum-sensing bacterial populations. *Nature* 2007;450(7168):411-414.

25. **Rumbaugh KP, Diggle SP, Watters CM, Ross-Gillespie A, Griffin AS et al.** Quorum sensing and the social evolution of bacterial virulence. *Curr Biol* 2009;19(4):341-345.

26. **Cornforth DM, Dees JL, Ibberson CB, Huse HK, Mathiesen IH et al.** *Pseudomonas aeruginosa* transcriptome during human infection. *Proc Natl Acad Sci U S A* 2018;115(22):E5125-E5134.

27. **D'Argenio DA, Wu M, Hoffman LR, Kulasekara HD, Deziel E et al.** Growth phenotypes of *Pseudomonas aeruginosa lasR* mutants adapted to the airways of cystic fibrosis patients. *Mol Microbiol* 2007;64(2):512-533.

28. **Mould DL, Botelho NJ, Hogan DA**. Intraspecies Signaling between Common Variants of *Pseudomonas aeruginosa* Increases Production of Quorum-Sensing-Controlled Virulence Factors. *mBio* 2020;11(4).

29. **Azimi S, Roberts AEL, Peng S, Weitz JS, McNally A et al.** Allelic polymorphism shapes community function in evolving *Pseudomonas aeruginosa* populations. *ISME J* 2020;14(8):1929-1942.

Figure 1

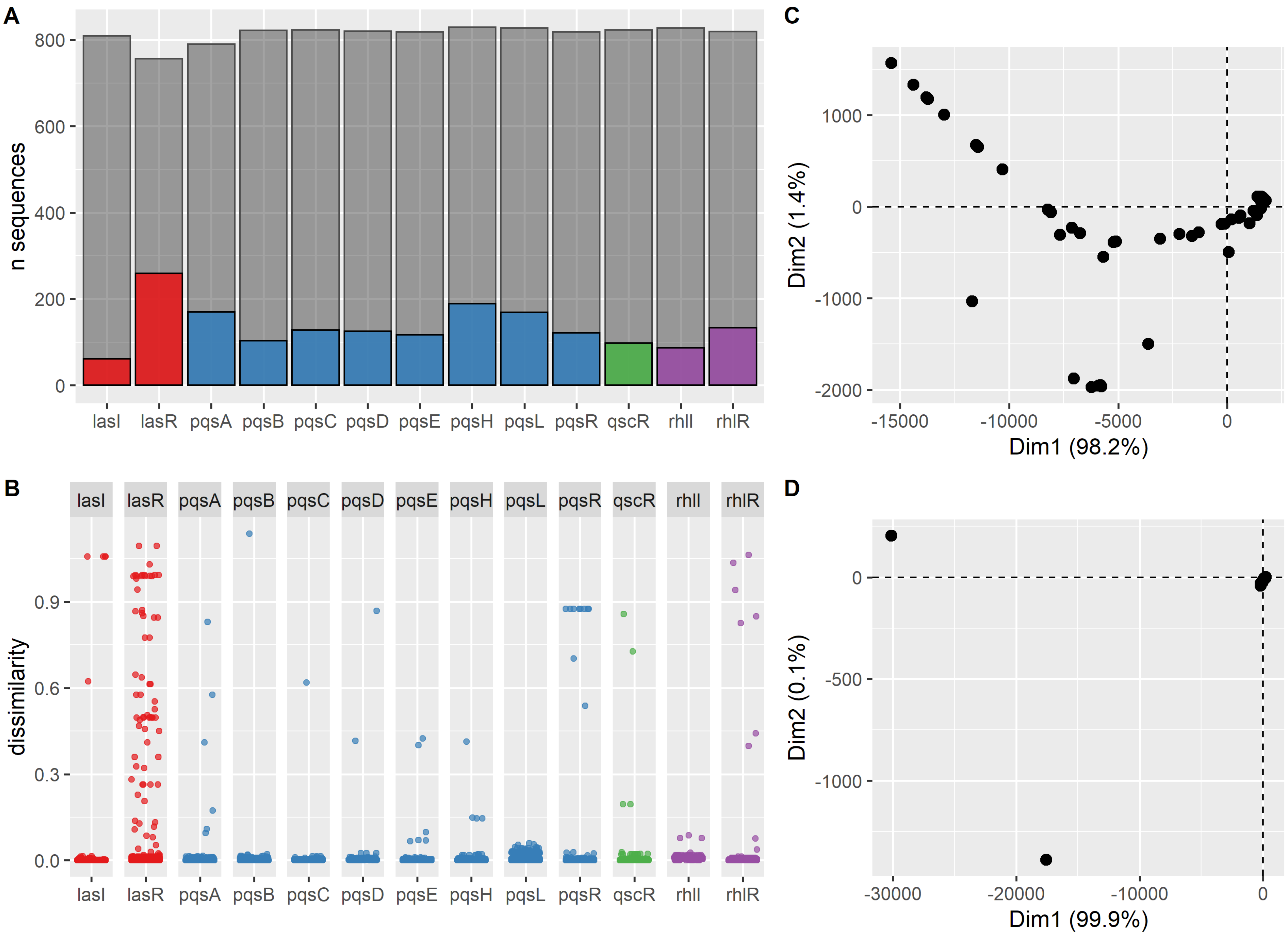
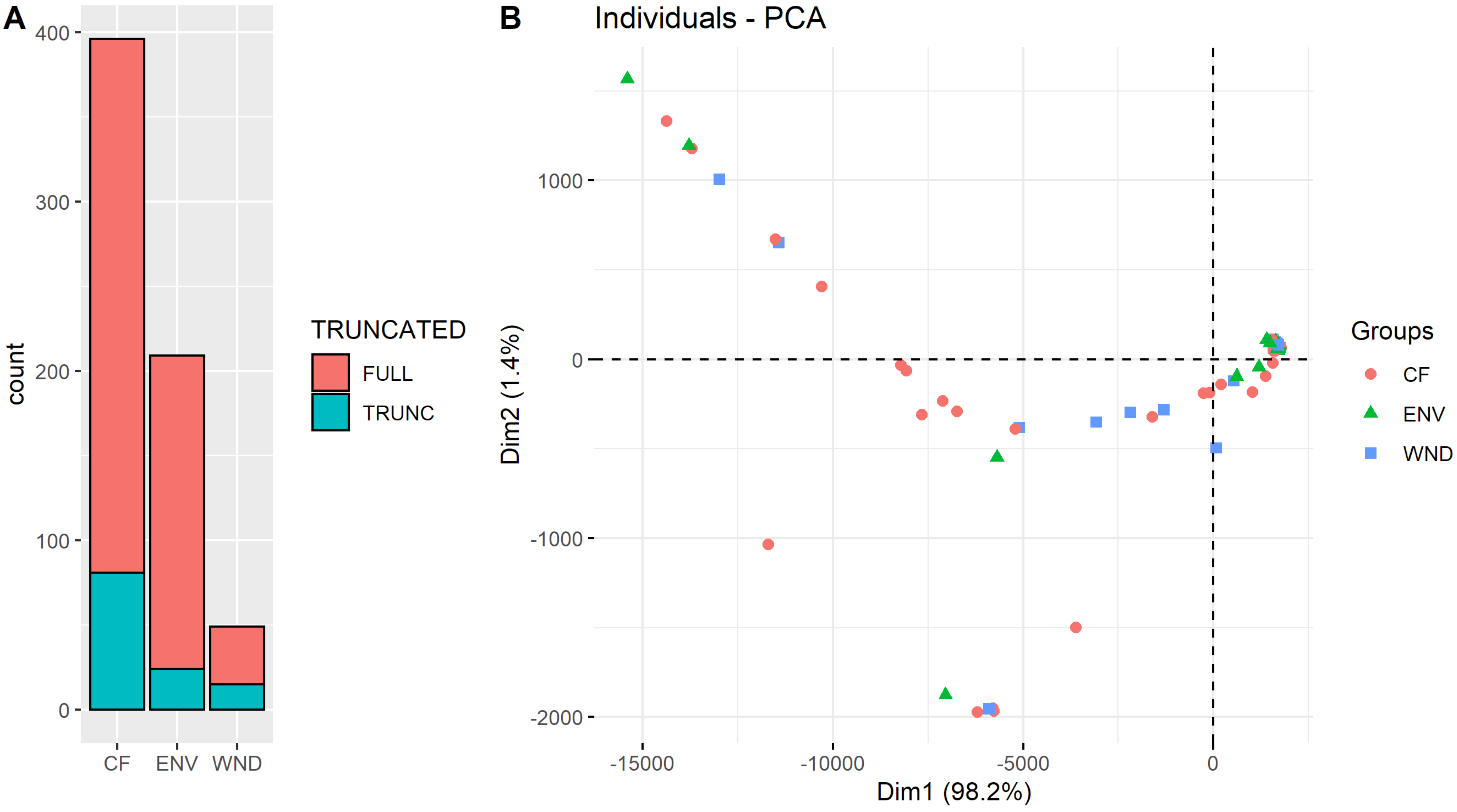

A
n sequences
800
600
400
200
0
lasI
lasR
pqsA
pqsB
pqsC
pqsD
pqsE
pqsH
pqsL
pqsR
qscR
rhlI
rhlR
C
Dim2 (1.4%)
1000
0
-1000
-2000
-15000
-10000
-5000
0
Dim1 (98.2%)
B
dissimilarity
lasI
lasR
pqsA
pqsB
pqsC
pqsD
pqsE
pqsH
pqsL
pqsR
qscR
rhlI
rhlR
0.9
0.6
0.3
0.0
lasI
lasR
pqsA
pqsB
pqsC
pqsD
pqsE
pqsH
pqsL
pqsR
qscR
rhlI
rhlR
D
Dim2 (0.1%)
0
-500
-1000
-30000
-20000
-10000
0
Dim1 (99.9%)

Figure 2


A
400
300
200
100
0
count
CF
ENV
WND
TRUNCATED
FULL
TRUNC
B
Individuals - PCA
1000
0
-1000
-2000
Dim2 (1.4%)
-15000
-10000
-5000
0
Dim1 (98.2%)
Groups
CF
ENV
WND