A Thesis

Entitled

Evolution of a global regulator – Lrp

by

Yvette O. Unoarumhi

Submitted to the Graduate Faculty as partial fulfillment of the requirements for the

Master of Science Degree in Biomedical Science

_________________________________________

Dr Jyl Matson, Committee Chair

_________________________________________

Dr Robert Blumenthal, Committee Member

_________________________________________

Dr Alexei Federov, Committee Member

_________________________________________

Dr Patricia R. Komuniecki, Dean
College of Graduate Studies

The University of Toledo

May 2016
An Abstract of
Evolution of a Global Regulator – Lrp

by

Yvette Ochuwa Unoarumhi

Submitted to the Graduate Faculty as partial fulfillment of the requirements for the
Master of Science Degree in Biomedical Sciences

The University of Toledo

May, 2016

Global regulators each control hundreds of genes in bacteria, and it is still unclear
how these regulators evolve, especially considering that gene regulation changes more
rapidly than the regulated genes themselves. Leucine-responsive regulatory protein (Lrp)
is a global regulator in enteric bacteria, controlling both metabolic and virulence-
associated genes. Lrp orthologs are found among both Bacteria and Archaea.
Surprisingly, even within the phylum γ-Proteobacteria, Lrp is a global regulator in some
orders and a local regulator in others. This raises important questions about the evolution
of Lrp functions. The way global regulators function is crucially important to bacterial
physiology. This thesis presents studies on the evolution and regulation pattern of Lrp,
carried out with the goal of providing insights into global regulators more generally.

Two independent studies of Lrp were carried out. The first compared Lrp
sequences from four bacterial orders within the γ-Proteobacteria: Enterobacteriales,
Vibrionales, Pasteurelales, and Alteromonadales. AsnC was also analyzed in parallel for
comparison, as it is a paralog of Lrp that in all known cases is a local regulator
controlling a small number of genes. As expected, Lrp and AsnC sequences formed two

iii
distinct clusters diverging from a common ancestor. These each divided into subclusters representing the Enterobacteriales, Vibrionales, and Pasteurellales. However, the Alteromonadales did not yield unitary clusters for either Lrp or AsnC, in contrast to the expected order-specific clustering we observed with the control housekeeping genes for 16S rRNA and RNA polymerase subunit RpoB. Logo analysis was also used to compare Lrp and AsnC in these four orders, and clear sequence signatures were identified. Ultimately, the Logo analysis provided the testable hypotheses that the globally-acting Lrp orthologs have short conserved sequences (particularly at the two ends of the polypeptides), and that Alteromonadales is unique among the orders tested in having member species with global or local Lrp orthologs.

The second study focused on the manner in which Lrp protein regulates expression of its own gene (lrp). In *E. coli* Lrp represses *lrp*. However, it has been reported that Lrp activates the *lrp* gene in *Vibrio cholerae*. This can have major consequences, since Lrp controls so many genes. To address this question we measured *lrp* expression in a different *V. cholerae* species, in the presence and absence of Lrp using a Plrp-lacZ transcriptional fusion. While the *V. cholerae* strain background and growth medium differ from the original study, the results indicate that Lrp represses *lrp* in *V. cholerae*, as in *E. coli*. Our studies of Lrp provide better understanding of global regulators, including testable hypothesis for future studies.
I dedicate this thesis to my mentor Gbenga Agboola and his family, my support. You all are one of the reasons this dream became a reality. Thank you for being a great mentor. I also dedicate this piece of work to all girls who aspire to be greater than the values their environment has placed on them.
Acknowledgements

I use this opportunity to thank my advisor Dr Jyl Matson for her immense contribution and guidance towards the successful completion of the thesis project. I appreciate your support and encouragement in my down times to propel me to success. Words are truly not enough to describe your contribution to my overall success as a scientist. I am indebted to Dr Robert Blumenthal for your unrelenting efforts towards molding me into a better scientist and person for future success. Thank you so much for all the many pep talks and guidance. Dr Alexei Federov, you motivated me to increase my knowledge of anything I find myself doing. You taught me all my programming skills and have provided great counsel on this project. I will also like to express my gratitude to Joanne Gray; you are indeed a rare gem, thank you so much for making my journey in graduate school easy. I thank my colleagues in the Matson Lab for their help.

I cannot do this without expressing my profound gratitude to my loving and ever supportive family. When I did not believe anything good was going to come out of my decisions, you all believed in me in and out of season. With you all standing by me, I know that I always had a support system. Love you now and always.

Lastly, I thank God, who has constantly kept me through my life journey. It has been You all the way and only You who makes me alive always and gives me the strength to carry on.
Table of Contents

Abstract ........................................................................................................................................ iii

Acknowledgments .................................................................................................................... vi

Table of Contents ..................................................................................................................... vii

List of Tables ............................................................................................................................... ix

List of Figures ............................................................................................................................. x

Chapter 1. Evolution of a Global Regulator: Lrp in Four Orders of γ-Proteobacteria ..... 1

1.1 Abstract ............................................................................................................................... 2

1.2 Background ......................................................................................................................... 4

1.3 Results & Discussion .......................................................................................................... 7

1.4 Conclusion .......................................................................................................................... 16

1.5 Methods .............................................................................................................................. 17

1.6 Figures ................................................................................................................................. 19

1.7 Additional File .................................................................................................................... 27
Chapter 2. Lrp autoregulation in Vibrio cholerae .............................................. 46

2.1 Abstract ........................................................................................................... 46
2.2 Introduction ....................................................................................................... 47
2.3 Materials & Methods ...................................................................................... 50
  2.3.1 Bacteria strains, media and growth ........................................................... 50
  2.3.2 β – galactosidase Assay ........................................................................... 51
2.4 Results ............................................................................................................... 52
2.5 Discussion .......................................................................................................... 54
2.6 References ......................................................................................................... 58
2.7 Figures ................................................................................................................. 64

3. Conclusion ........................................................................................................... 66

4. References .......................................................................................................... 69
List of Tables

1.1 Species used and accession numbers for their genomes and target genes ..........27
S1 Positions of E.coli Lrp residues highlighted in Figure 5 ..................................36
2.1 Strains used in this study.......................................................................................51
List of Figures

1.1 Role of Lrp superimposed on core genome phylogeny.......................... 19
1.2 Phylogeny and comparison of the paralog Lrp and AsnC ....................... 20
1.3 Comparison of Lrp orthologs grouped by order.................................. 22
1.4 Phylogeny of conserved housekeeping genes..................................... 24
1.5 Visualization of residues of interest in context of Lrp 3D structure........... 25
S1 Phylogenetic analysis of 40 paired Lrp and AsnC sequences from four orders of γ-Proteobacteria .......................................................... 31
S2 Comparison of orthologs from separate phylogenetic clusters in Alteromonadales .......................................................... 32
S3 Order- specific logo analysis of RecA ............................................... 33
S4 Single sequence alignment of divergent Alteromonadales Lrp or orthologs to sequences of Lrps playing known global or local roles.......................... 34
2.1 Expression of Lrp promoters in Vibrio cholerae .................................. 64
2.2 Conservation of lrp promoters in Vibrionaceae.................................... 65
Chapter 1

Evolution of a Global Regulator: Lrp in Four Orders of γ-Proteobacteria

Yvette Unoarumhi¹,², Robert M. Blumenthal¹,², and Jyl S. Matson¹*

¹Department of Medical Microbiology and Immunology, University of Toledo Medical School, Toledo, Ohio, USA

²Program in Bioinformatics and Proteomics/Genomics, University of Toledo Medical School, Toledo, Ohio, USA

* Corresponding author

Email: Jyl.Matson@utoledo.edu
1.1 ABSTRACT

**Background:** Bacterial global regulators each regulate the expression of several hundred genes. In *Escherichia coli*, the top seven global regulators together control over half of all genes. Leucine-responsive regulatory protein (Lrp) is one of these top seven global regulators. Lrp orthologs are very widely distributed, among both Bacteria and Archaea. Surprisingly, even within the phylum γ-Proteobacteria (which includes *E. coli*), Lrp is a global regulator in some orders and a local regulator in others. This raises questions about the evolution of Lrp and, more broadly, of global regulators.

**Results:** We examined Lrp sequences from four bacterial orders of the γ-Proteobacteria using phylogenetic and Logo analyses. The orders studied were Enterobacteriales and Vibrionales, in which Lrp plays a global role in tested species; Pasteurellales, in which Lrp is a local regulator in the tested species; and Alteromonadales, an order closely related to the other three but in which Lrp has not yet been studied. For comparison, we analyzed the Lrp paralog AsnC, which in all tested cases is a local regulator. The Lrp and AsnC phylogenetic clusters each divided, as expected, into subclusters representing the Enterobacteriales, Vibrionales, and Pasteurales. However, the Alteromonadales did not yield coherent clusters for either Lrp or AsnC. Logo analysis revealed signatures associated with globally- vs. locally- acting Lrp orthologs, providing testable hypotheses for which portions of Lrp are responsible for a global vs. local role. These candidate regions include both ends of the Lrp polypeptide but not, interestingly, the highly-conserved helix-turn-helix motif responsible for DNA sequence specificity.
Conclusions: Lrp and AsnC have conserved sequence signatures that allow their unambiguous annotation, at least in γ-Proteobacteria. Among Lrp orthologs, specific residues correlated with global vs. local regulatory roles, and can now be tested to determine which are functionally relevant and which simply reflect divergence. In the Alteromonadales, it appears that there are different subgroups of Lrp orthologs, one of which may act globally while the other may act locally. These results suggest experiments to improve our understanding of the evolution of bacterial global regulators.

Keywords: Transcription factors, Phylogenomics, Enterobacterales, Vibrionales, Pasteurellales, Alteromonadales
1.2 BACKGROUND

Global regulators (GRs) are transcription factors that, collectively, play a critical role in bacteria: they help to coordinate the responses of the cell’s thousands of genes to complex environmental changes [1]. In contrast to local regulators, which each control transcription of a small number of genes, GRs each control hundreds of genes. The top seven GRs in *Escherichia coli* (ArcA, Crp, Fis, Fnr, Ihf, Lrp, and NarL) together control about half of all its genes [2]. While each GR may have a general functional role, the genes controlled by each GR (its regulon) can specify a variety of disparate functions [2-5].

Despite their importance, a number of fundamental questions about GRs remain unanswered, in particular regarding the evolution of their global roles (see [6, 7]). Here, we use Lrp as a model GR to begin to address the question of GR evolution, focusing on the phylum and class that includes *E. coli* – the γ-Proteobacteria. This choice was made in part because, within different members of that phylum, there are examples of Lrp playing local and global roles. Further, this difference in Lrp role does not follow the same phylogenetic pattern as the core genome (Fig. 1, adapted from [8]). Specifically, Lrp appears to play global roles in many species of the order Enterobacteriales [9-16]; and in at least one [13] and possibly a second [17] species in the Vibrionales. In contrast, Lrp plays a local role (control of branched-chain amino acid biosynthesis) in the one tested species in the Pasteurellales, *Haemophilus influenzae* [18]. However, the Pasteurellales core genomes appear to be more closely related to those of Enterobacteriales than either is to the core genomes of the Vibrionales. While the relationship between these bacterial orders (and Fig. 1) is derived from analysis of concatenated gene sequences, and thus
have some level of uncertainty [19], it is nevertheless clear that Lrp plays different roles in closely-related bacterial orders, and for that reason is a good target for our studies on GR evolution.

Lrp has the functional flexibility one might expect of a GR. Lrp was originally named for its response to a coregulator (Leucine-responsive regulatory protein [20-22]), though subsequent analysis showed that it responds to a wider range of amino acids than just leucine [23]. Lrp was later recognized as belonging to a very large and ancient protein family (PF01037), with members in both the Bacteria and the Archaea [24-26]. This family is called the FFRPs, for Feast or Famine Regulatory Proteins, and the great majority includes two broad functional domains [27]. First, an amino-proximal helix-turn-helix DNA-binding domain, and second a coregulator response domain called RAM (Regulation of Amino Acid Metabolism) [27, 28]. The DNA sequence specificity of the Lrp helix-turn-helix is, in some cases, modulated by a flexible amino-terminal tail [29]. The RAM domain links coregulator levels to multimerization state, as follows. Lrp forms dimers that, in turn, tetramerize to form octameric rings with the helix-turn-helix domains exposed on the outer edge [30]. The DNA presumably wraps around this ring and, at least in the best-studied Lrp protein (from *Escherichia coli*; subsequently referred to as EcoLrp), apparently causes the octameric ring to open [31]. In the absence of coregulator, two EcoLrp octamers stack like coins to form a hexadecamer [32] and possibly larger complexes [33]. Leucine-RAM interactions favor dissociation of these 16mers back to two 8mers [34]. There is indirect evidence that the 16mers (low coregulator level) have higher affinity for DNA, while the 8mer (high coregulator level) has greater ability to activate transcription [35, 36]. Thus Lrp exhibits considerable regulatory flexibility – at
high-affinity operator sites on the DNA, the coregulator has little effect on repression and may increase the extent of activation (the 8mer remains bound but 16mer dissociation increases activation capacity); while at lower affinity operator sites the coregulator reduces both activation and repression.

To study the evolution of Lrp among γ-Proteobacteria, we focused on two questions. First, does the phylogeny of Lrp more closely follow its host’s core genome, or instead primarily reflect its global vs. local role? Second, are there any signature sequences associated with the global vs. local roles that might be used predictively during genome annotations? To address these questions, we examined the sequence changes in Lrp in four bacterial orders of the class γ-Proteobacteria. For comparison, we also studied a paralog of Lrp called AsnC, which consistently acts as a local regulator, in _E. coli_ controlling its own gene and the downstream _asnA_ gene (and another downstream gene post-transcriptionally) [37, 38]; as well as three housekeeping genes to reflect the core genomes (_rpoB_, _recA_, and _16S rRNA_).
1.3 RESULTS AND DISCUSSION

We examined the global regulator (GR) Lrp in the class \(\gamma\)-Proteobacteria, focusing on two orders in which Lrp acts globally (Enterobacteriales, Vibrionales), and one in which it acts locally (Pasteurellales; see Figure 1). In addition, we included one order in which the role of Lrp is currently unknown (Alteromonadales); this order is relatively closely related to the other three being studied and, like Vibrionales, includes many free-living marine bacteria. We included only species for which the genome sequence included orthologs for all of the genes we studied: \(lrp\), \(asnC\), 16S rRNA, \(rpoB\), and \(recA\) (Table 1).

Phylogeny and Identifying Motifs of the Paralogs Lrp and AsnC

We aligned the 80 amino acid sequences (40 Lrp and 40 AsnC, with both Lrp and AsnC sequences coming from the same genomes), and then subjected them to phylogenetic analysis (see Methods). The Lrp and AsnC sequences clustered separately, as shown in Figures 2A, 2B, and S1 (which shows the original joined Lrp/AsnC tree). This is not surprising, but requires a clarification. Namely, there were several cases of generic or mis-annotation associated with the genome sequences, where both genes were called “AsnC family” or something similar. We used logo analysis, which reveals patterns and extents of conservation within a set of orthologs [39, 40]. This analysis revealed both universally-conserved residues (within all Lrp + AsnC sequences), and residues that were highly conserved but distinct between Lrp and AsnC (indicated by shading in Figure 2C). These differences were then used to assign “AsnC family” polypeptides to the correct category. [Note that, unless otherwise specified, residue
numbers refer to the mass alignment positions, and these may differ from the numbering in the individual GenBank records.]

To assess the diagnostic value of these conserved sequence differences, we used the longest Lrp-specific segment (106-IQECHLVSGdFDYLLkTRV-124, where the two lower case symbols are not unique to Lrp; see Figure 2C) in a BLASTP search against the full nonredundant GenBank dataset. We examined the first 250 hits that had 100% query coverage and 100% identity. Of these, 64% were annotated as Lrp, 30% as “AsnC family”, 3% as “hypothetical protein”, and <1% each as “putative Lrp”, “transcriptional regulator” or “putative transcriptional regulator”. There were two cases, both in Vibrio genomes, annotated as the proline utilization regulator PutR. Significantly, there were no cases annotated as being AsnC. Conversely, when we used the equivalent segment from the AsnC sequence (VVEAYYTTG*YSIFk*M; * = wildcard), there were no cases annotated as “Lrp” – the great majority were labeled “transcriptional regulator”, with 8% annotated as “AsnC family” and 6% as AsnC. The sequence segments highlighted in Figure 2C may thus be useful in properly annotating Lrp and AsnC proteins, at least within the γ-Proteobacteria.

**Unusual Phylogenies Associated with Alteromonadales**

Closer examination of the Lrp and AsnC phylogenetic clusters reveals that the sequences cluster as expected by order for the Enterobacterales, Vibrionales and Pasteurellales (Figure 2 parts A and B). However, the Alteromonadales do not yield a single cluster for either protein (green shading), and this is true even when branches having <70% bootstrap support are collapsed (Fig. S1B). This is consistent with the order-specific logos we generated for Lrp (each derived from the 10 species used from
each order), shown in Figure 3A. There are a number of positions at which the Alteromonadales logo shows substantially lower conservation than in the other three orders. An example is in the carboxyl-proximal region (bottom of figure), positions 143-146, which is a strongly conserved GVND in three orders, and much more variable among the Alteromonadales. The differences between the two Alteromonadales clusters are shown, for both Lrp and AsnC, in Figure S2. We used two-sample Logo analysis [41], and the results reveal significant subcluster-specific sequence differences distributed over the entire length of the polypeptides. The subclusters thus reflect substantial sequence differences, not seen among Lrp or AsnC orthologs from the other three orders.

We considered the possibility that the core genomes for the Alteromonadales species we chose were inconsistently assigned. However, the phylogenies for two highly-conserved genes (16S rRNA, and RpoB – a large subunit of RNA polymerase) cluster as expected for all four orders (Figure 4, parts A and B). On the other hand, a third conserved gene – RecA – shows Alteromonadales-specific split clustering as was seen for Lrp and AsnC (Figure 4C). The order-specific logos for RecA, unlike the case for Lrp, do not reveal specific regions in which the Alteromonadales have unusual sequence variability (Figure S3).

Some bootstrap values in Fig. 1 are relatively low, particularly in the AsnC tree, but the separation of Paq, Plu, Ptu, and Isp Lrp orthologs from the other Alteromonadales Lrps is robust even when low-support nodes are collapsed (Fig. S1B). The separation of the two Alteromonadales RecA clusters also appears to be robust (Fig. 4C). Comparing the Alteromonadales Lrp, AsnC and RecA subclusters, there are some consistencies (Ffu/Fsp, Paq/Plu/Ptu, and Spe/Sfr/Slo are always together with one another) and some
differences (e.g., Mda and Isp have more variable associations). Detailed exploration of this phylogenetic pattern is beyond the scope of this study, but we note that similar disparities have been seen in some other studies that include Alteromonadales (e.g., MntX Mg++ transporter, Fig. S3a in [42]; various genes in [43]). This might reflect recent divergences or active horizontal gene transfer.

**Potential Differences between Globally- and Locally-acting Lrp Orthologs**

Even changing 1-2 amino acids in a transcription factor can significantly modify its regulatory activity [44, 45]. One of our major goals for this study was to identify sequence signatures that might be associated with global- vs. local-regulatory roles for Lrp. Accordingly, we used two-sample logo analysis [41] to compare the 20 presumed globally-acting Lrps (Enterobacteriales + Vibrionales) to the 10 presumed locally-acting Lrp orthologs (Pasteurellales) (Figure 3B). While bearing in mind the caveat that the number of genes controlled by Lrp has been tested directly in few of the 30 species included in this analysis, the residues identified by this analysis are testable candidate contributors to the global or local functionality of Lrp.

We consider the differing residues in four groups. First is the N-terminal 21 residues. This includes an N-terminal tail that plays a role in DNA binding [31] and sequence specificity (at least in Lrp from *E. coli*, *P. mirabilis*, and *V. cholerae*; [29]). The Pasteurellales Lrps have shorter and more variable N-termini. The two-sample Logo shows seven substantial differences in this region, including four differences over five residues, from positions 10-14.
Second is residues 36-60, which includes the DNA-recognizing helix-turn-helix (HTH) domain. Four major differences distinguish the globally- and locally-acting Lrp orthologs in this region. All four are relatively conservative, with one Glu/Asp difference, two Arg/Lys, and one Phe/Val. However, the D/E and one R/K change is within the first HTH helix, another R/K is within the recognition helix, and the F/V is three residues after the recognition helix. Between these and the differences in the N-terminal region, it is possible that sequence specificity differs between these two groups.

Third is residues 61-135, which includes the coregulator-binding RAM domain. There are nine residues with substantially-conserved differences between the global and local Lrp sets. None of the changes directly involve residues that form the coregulator-binding pocket (red arrows in Figure 3A). Three of the changes result in charge differences; two involve shifts from an aromatic (global) to a branched (local) sidechain (Tyr/Leu and Phe/Ile).

Finally, there are substantial differences at the C-termini, residues 159-171. At least in *E. coli* Lrp, this region is associated with changes in multimerization in response to the coregulator leucine [32]. In the Enterobacterales and Vibrionales, this is a highly-conserved LVIKTR motif, while in the Pasteurellales, only the K of that motif is (partially) conserved (Figure 3A). The two-source Logo shows three particularly significant conserved differences, of which the central one is starkest – Arg or Gln in the global Lrp set vs. Tyr or Phe in the local Lrp set (Figure 3B).

Figure 5 shows the distribution of these candidate role-specifying residues in the context of the Lrp three-dimensional structure. The figure shows four *E. coli* Lrp subunits (half of an octamer), with one subunit all in red to illustrate its overall shape, and another
subunit having candidate role-specifying residues as green spheres; as indicated in Figure 3, these are distributed over the full length of the protein (position numbers are given in Table S1). At least some of these apparent local vs. global differences, of course, may simply reflect genetic drift. But they represent a set of targets for specific functional testing in attempts to understand the differences between globally- and locally-regulating Lrp orthologs, and the more general question of what distinguishes these two classes of regulators.

**Lysine Acetylation**

Another potentially important level of control for GRs is post-translational modification. *E. coli* has enzymes that generate or remove acetyl groups from lysine residues [46]. While the role of Lrp acetylation has not been studied directly, a whole-proteome analysis of *E. coli* revealed that Lrp is substantially acetylated on three lysines: K28, K39 and K132 (supplementary data in [47]). These positions are indicated by blue arrows in Figure 3A (where the numbering reflects the multiple alignment), and orange spheres in Figure 5. K132 is less-well conserved in Vibrionales than in the other two orders, but is not strongly conserved in any of the orders. K39 is conserved in both the global and local Lrp sets, and is within the upstream helix of the HTH motif where acetylation might interfere with formation of a salt bridge to the DNA backbone, or even promote DNA binding [48]. Interestingly, K28 is strongly conserved in the Enterobacteriales and Vibrionales (global), but is replaced by Arg or His in the Pasteurellales (local), preserving the positive charge but not the acetylation potential. It
seems important to explore in future the possible role of Lrp acetylation, especially in bacteria where Lrp plays a global role.

**What is the likely Role of Lrp in the Alteromonadales?**

From the analyses presented in this study, it might be possible to make a testable predication as to the role (global or local) of Lrp in the Alteromonadales. From the phylogenetic relationships shown in Figure 2A, it seems possible that Lrp might play different roles in different species, corresponding to the distinct subclusters. However, the bootstrap values make it difficult to clearly assign any Lrp cluster as being particularly closely associated with the Pasteurellales (local role). Figure 3 suggests that at least the majority of Lrp orthologs in the Alteromonadales play a local role, based in particular on the missing or degenerate N-terminal and C-terminal regions. On the other hand, regarding some of the specific differences between local and global Lrp ortholog sets shown in Figure 3B, the Alteromonadales more closely resemble the global set. For example, in the Alteromonadales Lrp set Asp14 is more common than Ala14 (which we notate as D14 > A), along with N21 > K, E36 > D, R40 > K, F60 > V, F80 > V, S128 > A, and D/E136 > T. Only one of these positions, residue 21, differs substantially between the Alteromonadales Lrp subclusters (Figure S2).

These results are all ambiguous and make prediction difficult, but they are the result of comparing combined sequences. We therefore aligned *Moritella dasanensis* (Mda) Lrp individually to the known global regulator *E. coli* Lrp (Figure S4), based on Mda’s outlying position among the Enterobacteriales in the phylogenetic analysis shown in Figure 2A. These two Lrp orthologs share 91% identity, and it is particularly striking
that the conserved N- and C-terminal sequences characteristic of the global forms of Lrp are conserved in Mda, even though they are missing from most Alteromonadales Lrp orthologs. Also, 8/8 global signature residues (see preceding paragraph, underlined in Figure S4) are identical in Eco and Mda Lrp. It therefore seems reasonable to predict that Lrp will be found to play a global role in *M. dasanensis*. At the other extreme (Figure 2A) is the Lrp ortholog from *Idiomarina* spp. (Isp). It has just 68% identity to EcoLrp, comparable to the Isp identity with the known local regulator from *Haemophilus influenzae* (Hin), and matches Eco at just 2/8 signature residues. Thus it seems more likely that in Isp Lrp would be found to play a local regulatory role.

In contrast to the Alteromonadales, the Lrp orthologs we studied in the other three orders appear likely to play consistent roles – all local in Pasteurellales; all global in the Enterobacteriales and Vibrionales. Changes in bacterial regulatory networks, due in part to horizontal gene transfer, is well documented [7, 49]. It remains to be determined experimentally whether the proposed global/local role variation among Alteromonadales Lrp orthologs is real, but it raises questions about how the bacteria adapted to the gain or loss of a GR that would presumably have occurred during their evolution. Regarding loss, in *E. coli* deletion of the gene for Lrp does not greatly affect growth in rich media, but has profound effects under some conditions, and makes the cells far more sensitive to mutations affecting other regulators [50, 51]. Regarding displacement, in *E. coli* exchanging one Lrp ortholog for another (*Vibrio cholerae* or *Proteus mirabilis* for *E. coli*) results in only partial retention of the normal regulation of the several hundred genes in the Lrp regulon, despite their identical HTH motifs [13]. Introducing a new GR where none existed before would probably be the least disruptive of these scenarios, allowing
new genes to join the regulatory network over time. Presumably this latter gain-of-function scenario would result in substantially different regulon memberships than might be expected from simple species divergence, and this might provide additional evidence for past importation of a GR gene.
1.4 CONCLUSIONS

The global regulator Lrp, and its locally-acting paralog AsnC, have conserved sequence signatures that allow their unambiguous annotation, at least in γ-Proteobacteria. Among Lrp orthologs, we identified residues correlated with global vs. local regulatory roles, that can guide future experiments to determine which of them are functionally significant and which reflect simple divergence. Based on these observations, it was possible to make reasoned predictions for the global vs. local role of Lrp in the Alteromonadales, a bacterial order in which the role of Lrp has not yet been determined. Unlike the other three orders we studied here, it appears that in the Alteromonadales there are different subgroups of Lrp orthologs, one of which may act globally while the other may act locally. Together, these results suggest defined experimental avenues to improve our limited understanding of the evolution of global regulatory transcription factors in bacteria.
1.5 METHODS

Sequence Retrieval

Sequences were retrieved from the NCBI database (http://www.ncbi.nlm.nih.gov/). From each of the four orders we studied, we chose ten species having a known genome sequence that included orthologs for both Lrp and – for comparison – its locally-acting paralog AsnC and the core genome housekeeping genes 16S rRNA, \( rpoB \), and \( recA \). The species were chosen to, as far as possible, broadly represent the genera in each order. Table 1 shows the species used, along with accession numbers for the genome sequences and studied genes.

Phylogenetic Analyses

Multiple alignments of protein sequences were generated using MUSCLE and CLUSTALΩ [52, 53] with default parameters. Maximum likelihood phylogeny was constructed using the multiple sequence alignment results in FASTA format using the best parameters for the presented dataset by MEGA software (v6) (www.megasoftware.net/) [54]. Distance estimations were obtained by the pre-imputed JTT amino-acid substitution model [55] with 1000 bootstrap simulations. MEGA can use either the Dayhoff/PAM or JTT substitution matrices, and the JTT modeling was found to be optimal for the purpose of this study.

Logo Analyses

We used WebLogo (weblogo.berkeley.edu) to determine extent of conservation in aligned sequence sets [39, 40], and two-sample Logo (www.twosamplelogo.org) to compare two sets of aligned sequences [41].
1.6 Figures

Figure 1. Role of Lrp superimposed on core genome phylogeny. Five orders of the γ-Proteobacteria are shown, adapted (with permission) from a maximum likelihood tree generated by Gao et al. [8], and based on the concatenated sequences of 36 highly-conserved proteins. They used both maximum parsimony (MP) and maximum likelihood (ML) approaches, and the two numbers are the proportion of the puzzling quartets (ML) / % bootstrap scores (MP) that supported the given node. For each order, the colored shading and text to the right indicates the role played by Lrp in tested species (green = global, pink = local), and the tested species are also indicated. For two orders, indicated by “?” and yellow shading, the role of Lrp has not yet, to our knowledge, been tested.
Figure 2. Phylogeny and comparison of the paralogs Lrp and AsnC. Maximum likelihood phylogeny was constructed using the (A) Lrp and (B) AsnC protein sequences. The numbers above or below the internal branches show bootstrap values (%). Color keys indicate the different orders: magenta = Enterobacteriales (Ent), orange = Vibrionales (Vib), green = Alteromandales (Alt), red = Pasteurellales (Pas). (C) Logo comparison of all 40 Lrp vs. all 40 AsnC sequences. Areas visually identified as showing conserved differences are shaded in cyan.
Figure 3. Comparison of Lrp orthologs grouped by order. (A) The ten Lrp sequences from each order were used to generate aligned Logos, in order to compare globally- (Glb at right) and locally-acting (Loc) orthologs. The orders are abbreviated: Ent = Enterobacteriales, Vib = Vibrionales, Alt = Alteromonadales, Pas = Pasteurellales. The vertical arrows indicate positions of lysine acetylation (blue, from [47]) or formation of the coregulator binding pocket (red). See text for details. (B) Two-sample Logo comparing the global (Ent + Vib) and local (Pas) Lrp orthologs. Letters between the lines indicate amino acid residues that are conserved in both sets, symbols above the lines are selectively enriched in the globally-acting Lrp set, and symbols below the lines are selectively enriched in the locally-acting Lrp set.
Figure 4. Phylogeny of conserved housekeeping genes. Maximum likelihood phylogeny constructed for (A) 16S rRNA, (B) RpoB, and (C) RecA from the four bacterial orders. Colors are as assigned for Figure 1.
Figure 5. Visualization of residues of interest in context of Lrp 3D structure. The program VMD 1.9.2 was used to visualize half of an octameric ring of E. coli Lrp subunits (from PDB 2GQQ). VMD is developed with NIH support by the Theoretical and Computational Biophysics group at the Beckman Institute, University of Illinois at Urbana-Champaign. (A-D) are successive 90° rotations about the vertical axis. The topmost subunit has cyan spheres highlighting residues associated with Lrp-specific signatures (see Figure 2C), the next subunit is shown in red without additional highlighting, the next subunit shows in orange spheres the lysines that can be acetylated (see Figure 3A), and the bottom subunit shows in green spheres the residues associated with globally-acting Lrp orthologs (see Table S1 for position numbers of all highlighted residues).
<table>
<thead>
<tr>
<th>Species (abbr)</th>
<th>Genome</th>
<th>16s rRNA</th>
<th>RpoB</th>
<th>Ltp</th>
<th>AsnC</th>
<th>RecA</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Enterobacteriales</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Citrobacter freundii (Cfu)</td>
<td>GGT30259.1^a</td>
<td>AE237341</td>
<td>EFE53379.1</td>
<td>EFE55841.1</td>
<td>EFE6470.1</td>
<td>EFE6081.1</td>
</tr>
<tr>
<td>Dickeya dadantii (Dda)</td>
<td>CP002038.1</td>
<td>AF52077</td>
<td>WP_008900393.1</td>
<td>AQP98245.1</td>
<td>WP_00892115.1</td>
<td>WP_013318098.1</td>
</tr>
<tr>
<td>Escherichia coli (Eco)</td>
<td>NC_00913.3</td>
<td>NR_102804.1</td>
<td>NP_418414.1</td>
<td>NP_415409.1</td>
<td>NP_418199.1</td>
<td>AI290260.1</td>
</tr>
<tr>
<td>Klebsiella pneumoniae (Kpn)</td>
<td>APVW001000146.1^a</td>
<td>AF94537</td>
<td>CDQ52107.1</td>
<td>ECO213501.1</td>
<td>B5K3L3 KLEP3</td>
<td>KF75935.1</td>
</tr>
<tr>
<td>Photobacterium phosphoreum (Pae)</td>
<td>FM162511.1</td>
<td>NR_029093.1</td>
<td>CAQ82493.1</td>
<td>CAQ84933.1</td>
<td>KGM25901.1</td>
<td>CAQ85300.1</td>
</tr>
<tr>
<td>Proteus mirabilis (Pmi)</td>
<td>VY0407.1^a</td>
<td>KM095410.1</td>
<td>KGA91942.1</td>
<td>CAA71448.1</td>
<td>BAF4D6 PROMH</td>
<td>KGY5986.1</td>
</tr>
<tr>
<td>Salmonella enterica</td>
<td>NC_001397.1</td>
<td>DQ544537</td>
<td>AAF34993.1</td>
<td>NP_459935.1</td>
<td>NP_461775.1</td>
<td>NP_461750.1</td>
</tr>
<tr>
<td>Typhimurium (Sty)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Serratia marcescens (Sma)</td>
<td>UO2276.1^a</td>
<td>AF124635</td>
<td>KHO04269.1</td>
<td>AAA5466.1</td>
<td>KMU05309.1</td>
<td>WP_015670993.1</td>
</tr>
<tr>
<td>Yersinia enotella (Yen)</td>
<td>CP009567.1</td>
<td>NR_074538.1</td>
<td>AJU28227.1</td>
<td>AIF35937.1</td>
<td>AHM75528.1</td>
<td>AJU25926.1</td>
</tr>
<tr>
<td>Yersinia pestis (Ype)</td>
<td>AL590842.1</td>
<td>NR_074199.1</td>
<td>CAL22534.1</td>
<td>CAL20227.1</td>
<td>YP_00235098.1</td>
<td>CAL2198.1</td>
</tr>
<tr>
<td><strong>Vibrioionales</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grimontia indicia (Gin)</td>
<td>NZ_A9FND2000022.1^a</td>
<td>FJ943225</td>
<td>EOD77459.1</td>
<td>WP_002539165.1</td>
<td>WP_00254157.1</td>
<td>WP_00253758.1</td>
</tr>
<tr>
<td>Photobacterium profundum (Pap)</td>
<td>BMAM01000005.1^a</td>
<td>K74685</td>
<td>QSLW2.1</td>
<td>GAL04570.1</td>
<td>GAL07747.1</td>
<td>GAL08267.1</td>
</tr>
<tr>
<td><strong>Vibrio</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Vibrio anguillarum</strong> (Van)</td>
<td>CP000238.1</td>
<td>K73167</td>
<td>CDQ49218.1</td>
<td>AEB35539.1</td>
<td>CDQ51588.1</td>
<td>P26348.1</td>
</tr>
<tr>
<td><strong>Vibrio campbellii</strong> (Vca)</td>
<td>JSEV01000031.1^a</td>
<td>DQ808209</td>
<td>AYM361.1</td>
<td>KGB35110.1</td>
<td>KGB34501.1</td>
<td>WP_04534581.1</td>
</tr>
<tr>
<td><strong>Vibrio cholerae</strong> (Vch)</td>
<td>AE000852.1</td>
<td>AE001199</td>
<td>Q9V802.1</td>
<td>AAP3652.1</td>
<td>NP_229739.1</td>
<td>P45383.1</td>
</tr>
<tr>
<td><strong>Vibrio fischeri</strong> (Vfi)</td>
<td>NC_006479.2</td>
<td>CP0000170</td>
<td>WP_020579.1</td>
<td>WP_020427.1</td>
<td>WP_020464.1</td>
<td>WP_0203918.1</td>
</tr>
<tr>
<td><strong>Vibrio fluvialis</strong> (Vfl)</td>
<td>AWP701000031.1^a</td>
<td>HM771339</td>
<td>EGU40920.1</td>
<td>EGU40924.1</td>
<td>EGU40924.1</td>
<td>EGU40924.1</td>
</tr>
<tr>
<td><strong>Vibrio splendens</strong> (Vsp)</td>
<td>CM74173.1^a</td>
<td>AJU50301</td>
<td>WP_00324858.1</td>
<td>EAP92101.1</td>
<td>EAP92101.1</td>
<td>EAP92101.1</td>
</tr>
<tr>
<td><strong>Vibrio variabilis</strong> (Vva)</td>
<td>JNW011000005.1^a</td>
<td>GU019219</td>
<td>KU548154.1</td>
<td>KUA43484.1</td>
<td>GAL25642.1</td>
<td>GAL25642.1</td>
</tr>
<tr>
<td><strong>Vibrio vulnificus</strong> (Vvu)</td>
<td>BA000057.2</td>
<td>BA000057</td>
<td>KFC71115.1</td>
<td>BAC94085.1</td>
<td>NP_953861.1</td>
<td>AIL7194.1</td>
</tr>
</tbody>
</table>
### Pasteurellales

<table>
<thead>
<tr>
<th>Species</th>
<th>Accession Numbers</th>
<th>Genbank Accession Numbers</th>
<th>Type Accession Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Actinobacillus actinomycetemcomitans</em> (Ac)</td>
<td>CP000746.1</td>
<td>NR_074818.1</td>
<td>A6V7K5.1</td>
</tr>
<tr>
<td><em>Aeromonas hydrophila</em> (Aer)</td>
<td>CP003496</td>
<td>EKIX6954.1</td>
<td>WP_005540269.1</td>
</tr>
<tr>
<td><em>Anabaena sp.</em> (An)</td>
<td>NZ_CP007502.1</td>
<td>CP003496</td>
<td>EKIX6954.1</td>
</tr>
<tr>
<td><em>Caldobacterium leptum</em> (Ca)</td>
<td>NZ_AABP02000001.1*</td>
<td>AY488868</td>
<td>KKB02594.1</td>
</tr>
<tr>
<td><em>Campylobacter jejuni</em> (Cje)</td>
<td>NZ_CUM000012.1*</td>
<td>EU31064</td>
<td>KGQ369613.1</td>
</tr>
<tr>
<td><em>Diazotrophicus constructus</em> (Dco)</td>
<td>AC_000021.1</td>
<td>AF298641</td>
<td>WP_019516607.1</td>
</tr>
<tr>
<td><em>Eubacterium lentum</em> (El)</td>
<td>NC_002942.1</td>
<td>WZ6900</td>
<td>Q7W17.1</td>
</tr>
<tr>
<td><em>Franseria biformis</em> (Fbr)</td>
<td>JYK01000018.1</td>
<td>M55019</td>
<td>AJO91004.1</td>
</tr>
<tr>
<td><em>Granulicapsa pallida</em> (Gpa)</td>
<td>NZ_CP009571.1</td>
<td>NR_12882.1</td>
<td>AKA12987.1</td>
</tr>
<tr>
<td><em>Neisseria meningitidis</em> (Nme)</td>
<td>NZ_CCCM01000001.1*</td>
<td><em>NR_14428.1</em></td>
<td><em>WP_012093555.1</em></td>
</tr>
<tr>
<td><em>Pasturella multocida</em> (Pmu)</td>
<td>NZ_CP008918.1</td>
<td>NR_10816.1</td>
<td>AAA08211.1</td>
</tr>
</tbody>
</table>

### Alteromonadales

<table>
<thead>
<tr>
<th>Species</th>
<th>Accession Numbers</th>
<th>Genbank Accession Numbers</th>
<th>Type Accession Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Aeromonas hydrophila</em> (Aer)</td>
<td>AC_000021.1</td>
<td>AF298641</td>
<td>WP_019516607.1</td>
</tr>
<tr>
<td><em>Diazotrophicus constructus</em> (Dco)</td>
<td>AC_000021.1</td>
<td>AF298641</td>
<td>WP_019516607.1</td>
</tr>
<tr>
<td><em>Eubacterium lentum</em> (El)</td>
<td>NC_002942.1</td>
<td>WZ6900</td>
<td>Q7W17.1</td>
</tr>
<tr>
<td><em>Franseria biformis</em> (Fbr)</td>
<td>JYK01000018.1</td>
<td>M55019</td>
<td>AJO91004.1</td>
</tr>
<tr>
<td><em>Granulicapsa pallida</em> (Gpa)</td>
<td>NZ_CP009571.1</td>
<td>NR_12882.1</td>
<td>AKA12987.1</td>
</tr>
<tr>
<td><em>Neisseria meningitidis</em> (Nme)</td>
<td>NZ_CCCM01000001.1*</td>
<td><em>NR_14428.1</em></td>
<td><em>WP_012093555.1</em></td>
</tr>
<tr>
<td><em>Pasturella multocida</em> (Pmu)</td>
<td>NZ_CP008918.1</td>
<td>NR_10816.1</td>
<td>AAA08211.1</td>
</tr>
</tbody>
</table>

*contigs and/or full genome were not available at time of writing*
1.7 ADDITIONAL FILES

Additional file 1: Figures S1-S4, Table S1. The figures show a universal phylogeny of 80 Lrp and AsnC sequences, two-source Logos comparing Alteromonadales subclusters for Lrp and AsnC, order-specific Logos for RecA, and alignments of selected individual Lrp sequences. The table shows residues chosen for highlighting in the structural representation of Lrp shown in Figure 5.

1.7.1 ABBREVIATIONS

FFRP: Feast-or-famine regulatory protein; GR: global regulatory transcription factor; HTH: helix-turn-helix DNA-binding motif; Lrp: Leucine-responsive regulatory protein; RAM: regulation of amino acid metabolism motif. Abbreviations for the 40 bacterial species studied are shown in Table 1.

1.7.2 COMPETING INTERESTS

The authors declare that they have no competing interests.

1.7.3 AUTHORS’ CONTRIBUTIONS

JM and RB designed the study and helped to draft the manuscript. YU performed the work and wrote the manuscript. All authors edited and approved the final manuscript.
1.7.4 ACKNOWLEDGEMENTS

We are grateful to Dr. R.S. Gupta (McMaster University) for permission to adapt a figure from one of his publications, and Dr. I.S. Novella (University of Toledo) for critically reading the manuscript. This work was supported by startup funds from the University of Toledo College of Medicine and Life Sciences (JSM). YU was supported in part by a stipend and tuition scholarship from the UT College of Graduate Studies.
1.7.5 Supplementary Figures

Figure S1. Phylogenetic analysis of 40 paired Lrp and AsnC sequences from four orders of \( \gamma \)-Proteobacteria.
Maximum likelihood phylogeny was constructed using Lrp and AsnC protein sequences. The numbers above or below the internal branches show bootstrap values (%). Color keys indicate the different orders: magenta = Enterobacterales (Ent), orange = Vibrionales (Vib), green = Alteromonadales (Alt), red = Pasteurellales (Pas). A – Full tree. B – Tree with nodes collapsed if they had <70% bootstrap support.
Figure S2. Comparison of orthologs from separate phylogenetic clusters in Alteromonadales. Two-sample Logo analysis was used to compare the orthologs of Lrp (A) or AsnC (B) from separate phylogenetic clusters (see Figure S1). The species abbreviations are listed in Table 1.

Alteromonadales. Two-sample Logo analysis was used to compare the orthologs of Lrp (A) or AsnC (B) from separate phylogenetic clusters (see Figure S1). The species abbreviations are listed in Table 1.
Figure S3. Order-specific Logo analysis of RecA. RecA sequences were taken from the same genomes as were used to obtain Lrp and AsnC sequences, and separated by order (Ent = Enterobacteriales, Vib = Vibrionales, Alt = Alteromonadales, and Pas = Pasteurellales). Numbering refers to positions in the alignment of all 40 sequences.
Figure S4. Single-sequence alignments of divergent Alteromonadales Lrp orthologs to sequences of Lrps playing known global or local roles. (A) Alignment of Moritella

<table>
<thead>
<tr>
<th></th>
<th>Score</th>
<th>Expect</th>
<th>Identities</th>
<th>Positives</th>
<th>Gaps</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>303</td>
<td>2e-102</td>
<td>149/164(91%)</td>
<td>156/164(95%)</td>
<td>0/164(0%)</td>
</tr>
</tbody>
</table>

Mda 1  
MKEVTRPMKELDRNILNLEKQGRISNLVELSKVRLVSPCPLERLDRVRLQIYG 60  
M++ K RP M+K LDRNILEKQGRISNLVELSKVRLVSPCPLERLDRVRLQIYG

Eco 1  
MKEVTRPMKELDRNILNLEKQGRISNLVELSKVRLVSPCPLERLDRVRLQIYG 60  
M+K LDRNILEKQGRISNLVELSKVRLVSPCPLERLDRVRLQIYG

Mda 61  
YTAIMQFQLSSLSLVEITLNRAGADVFQFQNFKAVQEIEEEIEQCHLVSDFDY LLKTR 120  
YTAIMQFQLSSLSLVEITLNRAGADVFQFQNFKAVQEIEEEIEQCHLVSDFDY LLKTR

Eco 61  
YTAIMQFQLSSLSLVEITLNRAGADVFQFQNFKAVQEIEEEIEQCHLVSDFDY LLKTR 120  
YTAIMQFQLSSLSLVEITLNRAGADVFQFQNFKAVQEIEEEIEQCHLVSDFDY LLKTR

Mda 121  
VCDMSAYRLLLGETLLRLPGVNDTRTYVMEEVQKSNRLVIKTR 164  
V CDMSAYRLLLGETLLRLPGVNDTRTYVMEEVQKSNRLVIKTR

Eco 121  
VCDMSAYRKLGETLLRLPGVNDTRTYVMEEVQKSNRLVIKTR 164  
VCDMSAYRKLGETLLRLPGVNDTRTYVMEEVQKSNRLVIKTR

B. Score | Expect | Identities | Positives | Gaps |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>213</td>
<td>6e-67</td>
<td>102/151(68%)</td>
<td>125/151(82%)</td>
<td>0/151(0%)</td>
</tr>
</tbody>
</table>

Isp 9  
IDLRDKILRLQERGRISNLKVLAKQGLPSLPCPLERTVRKLERNGVYHARIQDKL 68  
IDLRDKILRLQERGRISNLKVLAKQGLPSLPCPLERTVRKLERNGVYHARIQDKL

Eco 12  
IDLRDKILRLQERGRISNLKVLAKQGLPSLPCPLERTVRKLERNGVYHARIQDKL 71  
IDLRDKILRLQERGRISNLKVLAKQGLPSLPCPLERTVRKLERNGVYHARIQDKL

Isp 69  
TLQFVLVEITLSTSSDFAESFAARKEITEFEECHLVSDFDLLKARVADMAAYRKLL 128  
TLQFVLVEITLSTSSDFAESFAARKEITEFEECHLVSDFDLLKARVADMAAYRKLL

Eco 72  
ASLQFVLVEITLSTSSDFAESFAARKEITEFEECHLVSDFDLLKARVADMAAYRKLL 131  
ASLQFVLVEITLSTSSDFAESFAARKEITEFEECHLVSDFDLLKARVADMAAYRKLL

Isp 129  
GETLLNMGVPVNSRYTEYVMEAVKQENKVVIK 159  
GETLLNMGVPVNSRYTEYVMEAVKQENKVVIK

Eco 132  
GETLLNMGVPVNSRYTEYVMEAVKQENKVVIK 162  
GETLLNMGVPVNSRYTEYVMEAVKQENKVVIK

C. Score | Expect | Identities | Positives | Gaps |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>201</td>
<td>2e-65</td>
<td>98/152(64%)</td>
<td>120/152(78%)</td>
<td>0/152(0%)</td>
</tr>
</tbody>
</table>

Isp 8  
TVDLRDKKILRLQERGRISNLKVLAKQGLPSLPCPLERTVRKLERNGVYHARIDPKL 67  
TVDLRDKKILRLQERGRISNLKVLAKQGLPSLPCPLERTVRKLERNGVYHARIDPKL

Hin 8  
TVDLRDKKILRLQERGRISNLKVLAKQGLPSLPCPLERTVRKLERNGVYHARIDPKL 67  
TVDLRDKKILRLQERGRISNLKVLAKQGLPSLPCPLERTVRKLERNGVYHARIDPKL

Isp 68  
PVLQFVLVEITLSTSSDFAESFAARKEITEFEECHLVSDFDLLKARVADMAAYRKLL 127  
PVLQFVLVEITLSTSSDFAESFAARKEITEFEECHLVSDFDLLKARVADMAAYRKLL

Hin 68  
PVLQFVLVEITLSTSSDFAESFAARKEITEFEECHLVSDFDLLKARVADMAAYRKLL 127  
PVLQFVLVEITLSTSSDFAESFAARKEITEFEECHLVSDFDLLKARVADMAAYRKLL

Isp 128  
LGTTINMGVPVNSRYTEYVMEAVKQENKVVIK 159  
LGTTINMGVPVNSRYTEYVMEAVKQENKVVIK

Hin 128  
LGTTINMGVPVNSRYTEYVMEAVKQENKVVIK 159  
LGTTINMGVPVNSRYTEYVMEAVKQENKVVIK
dasanensis (Mda) to E. coli (Eco, global) Lrp. (B) Alignment of Idiomarina spp. (Isp) to Eco Lrp. (C) Alignment of Isp to Haemophilus influenzae (Hin, local) Lrp.
<table>
<thead>
<tr>
<th>Residues(^a) distinguishing all(Lrps) from all(AsnCs(____)</th>
<th>Residues(^b) distinguishing all(Global) from all(Local)Lrps</th>
</tr>
</thead>
<tbody>
<tr>
<td>ILE 29</td>
<td>ARG 14</td>
</tr>
<tr>
<td>CYM 30</td>
<td>LYS 25</td>
</tr>
<tr>
<td>ASN 31</td>
<td>ASP 26</td>
</tr>
<tr>
<td>VAL 32</td>
<td>ARG 28</td>
</tr>
<tr>
<td>THR 43</td>
<td>ARG 54</td>
</tr>
<tr>
<td>PRO 44</td>
<td>TYR 69</td>
</tr>
<tr>
<td>CYS 45</td>
<td>LEU 116</td>
</tr>
<tr>
<td>LEU 46</td>
<td>ARG 73</td>
</tr>
<tr>
<td>GLN 47</td>
<td>PHE 77</td>
</tr>
<tr>
<td>ARG 50</td>
<td>ASP 83</td>
</tr>
<tr>
<td>LEU 52</td>
<td>ALA 86</td>
</tr>
<tr>
<td>GLN 53</td>
<td>GLN 92</td>
</tr>
<tr>
<td>ARG 54</td>
<td>ILE 161</td>
</tr>
<tr>
<td>ASP 71</td>
<td>THR 163</td>
</tr>
<tr>
<td>MET 124</td>
<td>ARG 164</td>
</tr>
<tr>
<td>ALA 72</td>
<td>GLY 140</td>
</tr>
<tr>
<td>SER 73</td>
<td>ASN 142</td>
</tr>
<tr>
<td>LEU 74</td>
<td>ASP 143</td>
</tr>
<tr>
<td>LEU 75</td>
<td>ASP 88</td>
</tr>
<tr>
<td>VAL 76</td>
<td>VAL 89</td>
</tr>
<tr>
<td>GLU 79</td>
<td>PHE 109</td>
</tr>
<tr>
<td>PRO 87</td>
<td>MET 150</td>
</tr>
<tr>
<td>ASP 88</td>
<td>GLU 151</td>
</tr>
<tr>
<td>VAL 89</td>
<td>VAL 152</td>
</tr>
<tr>
<td>PHE 90</td>
<td>LYS 25</td>
</tr>
<tr>
<td>VAL 97</td>
<td>LYS 36</td>
</tr>
<tr>
<td>PHE 93</td>
<td>LYS 129</td>
</tr>
<tr>
<td>ASN 94</td>
<td>ALA 96</td>
</tr>
<tr>
<td>VAL 97</td>
<td>VAL 153</td>
</tr>
<tr>
<td>ALA 96</td>
<td>LYS 154</td>
</tr>
</tbody>
</table>

Table S1. Positions of *E. coli* Lrp residues highlighted in Figure 5

\(^a\) Based on the 40 species included in this study, with ten each from the Enterobacteriales, Vibrionales, Alteromonadales and Pasteurellales. Numbering is

35
from PDB 2GQQ (E. coli Lrp), and differs from numbering in the rest of the figures (which is based on the multiple alignment).

b – Assuming that all ten Enterobacteriales and Vibrionales Lrps used are global, all ten Pasteurellales are local, and excluding the Alteromonadales. This set derived from the order-specific Logos in Figure 3A.

c – Taken from supplementary information of reference 47.
1.8 REFERENCES


Chapter 2

Lrp Autoregulation in *Vibrio cholerae*
2.1 Abstract

Leucine responsive regulatory protein (Lrp) is a global regulator that controls hundreds of genes in a range of Bacteria. It can regulate genes negatively or positively. It is still a puzzle how Lrp functions as global or local regulator in different bacteria orders. This work is focused on regulation of the \textit{lrp} promoter in strains of the gram negative pathogen \textit{Vibrio cholerae}. Previously, it was shown that Lrp activates its own gene (\textit{lrp}) in a native \textit{V. cholerae} background. We used \textit{lrp}-lacZ transcriptional fusion from that study and transformed into both WT and \textit{lrp} insertion mutants of the commonly - used \textit{V. cholerae} strains C6706 and A1552. Relative LacZ activity was measured in each strain. In comparison to the wild type strain A1552, A1552\textit{Δlrp} showed as previously found decreased expression, indicating activation. However, C6706 \textit{Δlrp} showed increased expression compared to wild type C6706. These results suggest that the regulation of \textit{lrp} involves Lrp acting as a repressor in at least one \textit{V. cholerae} background.
2.2 Introduction

Lrp is one of the top seven global regulators in E. coli [1]. Together the seven control over half of all the genes in E. coli K-12; Lrp itself controls about 200 genes directly, and about the same number indirectly [2-4]. E. coli Lrp (EcoLrp) affects a variety of genes, including several associated with pathogenicity in non-K-12 strains [2, 5, 6], but plays particularly important roles in nitrogen and amino acid metabolism [5, 7-10]. Very broadly, the role of Lrp in E. coli appears to be helping the cell adjust to changes between nutrient-rich and nutrient-limited environments (“feast or famine”; [5]) or, as some have put it, “between the gut and the gutter” [11].

While there have been extensive studies on Lrp in E. coli, and to some extent in other members of the order Enterobacteriales, very little work has been carried out in the marine order Vibrionales, which includes the significant pathogen Vibrio cholerae. Lrp is a global regulator in V. cholerae [12] and it shares an orthologous relationship with E. coli Lrp [13] A sequence comparison of E. coli and V. cholerae Lrp orthologs is presented within chapter 1, and so will not be discussed here, but overall they share 92% amino acid sequence identity, including 100% identity of their DNA-recognizing helix-turn-helix domains. Thus, one might expect Lrp to play a major regulatory role in V. cholerae and other Vibrionales, making it a worthy target of study.

In comparison to other global regulators such as Crp (cAMP receptor protein), the role of Lrp as a global regulator has not been well studied in V. cholerae. A PubMed search (February, 2016) found just six publications with a focus on the regulatory role of Lrp in Vibrio: one showing that the quorum sensing regulator HapR can act by
antagonizing Lrp [12]; another showing that the V. cholerae Lrp can function in E. coli (PMID [14]); a trio about V. vulnificus Lrp showing importance to stress resistance, Lrp control of the expression of an acid-resistance cadaverine export system, and Lrp control of a proton transporter [15-17] and influence of lrp on survival in stressful conditions [17] and one on the role of the N terminus of V. cholerae Lrp [18]. An unpublished study, by the laboratories of Robert Blumenthal, Arkady Khoudursky, and Gary Schoolnik, assessed the role of Lrp in V. cholerae using microarrays and competition assays in rabbit ileal loops, comparing a Lrp± V. cholerae strain pair (R.Blumenthal, personal communication). Results from these experiments showed that 1) Lrp affects competition and fluid accumulation in ileal loop assays, 2) Lrp affects expression of a significant number of genes both in vivo and in vitro through direct or indirect regulation. Together, these findings clearly suggest that Lrp is playing an important role in V. cholerae.

Lrp consists of an N–terminal domain with a helix–turn–helix motif that binds DNA, and a C–terminal RAM domain (Regulation of Amino acid Metabolism) that responds to amino acid coregulators. Initially, it was discovered that Lrp (in E.coli) responds to a limited range of coregulators such as leucine [19-24] and alanine [25-30], though there was no systematic test of other possible coregulators. Subsequently, Lrp was shown to respond to a broad range of coregulators including many other amino acids [31]. These coregulators are thought to mediate formation of Lrp multimers, shifting the octamer-hexadecamer equilibrium towards the octamers ([19, 21].

Another factor in the activity of any regulator is its amount in the cell, and V. cholerae Lrp may be regulated in a different way than in E.coli. The regulation of Lrp was studied by Lintner et al. [14], and revealed that V. cholerae resembled E. coli and
Proteus mirabilis in exhibiting reduced levels in rich medium (compared to glucose minimal medium)

Those authors also examined the ability of Lrp to regulate its own expression in different bacteria. To this end, the promoter regions of Lrp from E. coli, Proteus mirabilis and V. cholerae were fused to a promoterless lacZ gene. The lacZ fusions were introduced into wildtype and lrp deletion strain pairs of E. coli and V. cholerae. In E. coli, Lrp represses its own promoter, Plrp [32]. In contrast to the repression observed for EcoLrp (and the other Enterobacteriales member P. mirabilis), Lrp appeared to be activating Plrp in V. cholerae [14]. As the levels of a major regulatory protein can have substantial effects on cell physiology, it was important to test whether this surprising result was generally true of V. cholerae strains. Interestingly, when the Plrp-lacZ fusion from V. cholerae was moved into an E. coli or P. mirabilis background, it was strongly repressed by the two heterologous Lrp proteins. Together, these results suggested a more complex regulation of Lrp in V. cholerae than in the two Enterobacteriales species.

In this chapter, using the discoveries from the Lintner paper as a foundation for my work, we explored the auto regulatory role of Lrp in an additional strain of V. cholerae. We hoped to determine if the previously observed regulatory pattern of Lrp was unique to the strain studied or if it was conserved in other strains of V. cholerae. While both strains used here are El Tor isolates of V. cholerae, the one used by Lintner et al. (A1552) is studied in a relatively limited number of laboratories, while the C6706 strain is much more widely used.
2. 3 Materials & Methods

2.3.1 Bacteria strains, Media and Growth Conditions

The bacterial strains used for this study are listed in Table 1. All strains were maintained at -80°C in Luria-Bertani (LB) broth (ref) supplemented with 20% glycerol. The antibiotic used, where indicated, is ampicillin was used at 100 µg/ml to select for plasmid pPM3006. Plasmid pPM3006 was purified from strain A1552 using the Qiagen Spin Miniprep Kit. Plasmids were introduced into *V. cholerae* strains by electroporation.

2.3.2 β-Galactosidase Assay

Strains were grown overnight at 37°C in LB medium containing ampicillin. A 100 µl portion of the overnight cultures was subcultured into 5 ml of fresh medium and grown for 3 h. One ml of culture was added to 4 µl of 50 mg/µl chloramphenicol at 0° and incubated for 20 minutes to stop growth. The optical density at 600nm of another 1 ml of the culture was measured using a spectrophotometer. Chloramphenicol-treated cells (100 µl) and 900µl Z buffer (ref OR recipe) were mixed on ice. Cells were lysed by adding 20 µl 0.1% SDS and 40µl chloroform. Lysed cells were vortex mixed for 10 s and placed in a 37 °C incubator for 10-15 min.
### Table 1: Strains used in this Study

<table>
<thead>
<tr>
<th>Strain</th>
<th>Description</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Vibrio cholerae El tor A1552</em></td>
<td>WT</td>
<td>G. K Schoolnik</td>
</tr>
<tr>
<td><em>Vibrio cholerae El tor A1552 lrp</em></td>
<td><em>lrp-cat</em> derivative of strain A1552</td>
<td>G.K. Schoolnik</td>
</tr>
<tr>
<td><em>Vibrio cholerae El tor C6706</em></td>
<td>WT</td>
<td>J. Matson</td>
</tr>
<tr>
<td><em>Vibrio cholerae El tor C6706 lrp</em></td>
<td><em>lrp : :Tn10</em></td>
<td>J. Matson</td>
</tr>
</tbody>
</table>

Levels of β-galactosidase were determined by adding 200 µl 4mg/ml o-nitrophenyl-D-galactosidase substrate (ONPG in Z buffer) to lysed cells, and vortex mixed for 5 s. The reaction was stopped by adding 500 µl Na$_2$CO$_3$ when yellow color was observed. The optical density at 420nm and 550nm were measured. β-galactosidase activity was calculated using the Miller equation:

$$\text{Miller Units} = \frac{(\text{OD}_{420} - (1.75 \times \text{OD}_{550})) \times 1000}{(\text{OD}_{600} \times T \times V)}$$

Where T is the time between ONPG and Na$_2$CO$_3$ additions, and V is the volume of lysed cells used.
2.4. RESULTS

Lrp expression in three bacteria (*V. cholerae*, *E. coli*, and *P. mirabilis*) was previously studied by looking at activity of the three different *lrp* promoters in wild type and *lrp*-null strains of each species [14]. Unlike in *E. coli* and *P. mirabilis*, where Lrp appears to repress its own expression, the *lrp* promoter in *V. cholerae* showed a higher expression level in the presence of Lrp, suggesting autoactivation [14]. For this study, we wanted to determine if the previously observed regulatory pattern of Lrp was unique to the one strain previously studied or if it was conserved in other strains of *V. cholerae*. Therefore, we repeated the previous studies using strain A1552 and adding the more commonly used *V. cholerae* strain C6706 for comparison.

The previously constructed *Plrp*:lacZ transcriptional fusion plasmid pPM3005 [14] was introduced into strain C6706 and a *lrp* insertion mutant derivative through transformation. The *lrp* mutant of C6706 was obtained from the nonredundant *V. cholerae* transposon insertion library [33]. Lrp promoter activity was assessed in A1552, C6706, and the corresponding *lrp* insertion mutants from each background using a β-galactosidase assay as described in the materials and methods.

We observed very low expression from *Plrp* promoter in the presence or absence of Lrp in strain A1552 (Figure 1). This was probably due in part to our use of a rich medium (LB), as at least in *E. coli* *lrp* transcription is low in rich media [32]. However, expression was somewhat higher in the Lrp*"* background, as was observed previously [14], though results from both strains are not significantly different in both Lrp*"* strains. Whether or not there is evidence of activation by Lrp, there is clearly no evidence for repression. In C6706, we again saw very low activity of the *lrp* promoter in the Lrp*"*
background. However, we saw much higher activity in the absence of Lrp, which is very different from other results. The higher in C6706 lrp background, is an exception in our findings. This finding is similar to the previously observed activity of the *V. cholerae lrp* promoter in the presence and absence of Lrp in the heterologous backgrounds of *E. coli* and *Proteus mirabilis* [14]. These results suggest that there are strain-specific differences in *lrp* regulation in *V. cholerae*. 
2.5 Discussion

In this study, we examined the regulatory functions of a top seven global regulator in *E.coli* -Lrp in the Vibrionales order, *Vibrio cholerae*. Lrp has been used as a model for understanding roles of global regulators and its evolution. Previously, mechanisms of Lrp regulation were mostly inferred from results in the well-studied model bacterium *E. coli*. In addition to the information known about the role of *Vibrio cholerae* Lrp (VchLrp), very little is known about its regulatory role. In *E. coli*, Lrp directly regulates hundreds of genes and indirectly influences the expression of hundreds more ([5]. Unpublished results (R.M Blumenthal, pers. commun.) indicate that Lrp also controls a large number of genes in *V. cholerae*.

If a protein controls hundreds of genes, then regulation of its own gene’s expression can have a large impact on the cell’s physiology. In *E.coli*, the global regulator H-NS has been reported to directly control *P*_{lrp} [34], and the nitric oxide-responsive regulator NsrR binds in the *lrp* promoter region [35], though its role in controlling *lrp* has not yet been confirmed. Lintner et al. studied the Lrp feedback repression on its promoter *P*_{lrp} in three organisms: *E. coli*, an additional member of the Enterobacteriales *Proteus mirabilis*, and a member of the Vibrionales *V. cholerae*. It was found that in *E. coli* and *P. mirabilis*, as expected, *lrp* transcription was higher in a Lrp\(^{-}\) than in a Lrp\(^{+}\) background. However, the reverse was true in *V. cholerae* – *lrp* transcription was elevated in the presence of VchLrp. This effect was dependent on both the promoter and the background: *P*_{lrp} from *Vibrio* was repressed by Lrp in *E. coli* and *P. mirabilis*, while *P*_{lrp} from *E. coli* and *P. mirabilis* was repressed by Lrp in *V. cholerae*. 
Given the apparently major regulatory role of Lrp in *V. cholerae*, it was important to explore the reported difference in control of *lrp* in the *Vibrio* background. This is significant both for what can be learned about *V. cholerae* physiology, and also for insight into how much regulation in one organism can be extrapolated into regulation in another based on sequence similarities alone. In particular, I sought to determine if the *V. cholerae* Lrp feedback activation was unique to the particular strain used or a conserved feature in Vibrionaceae order, and found that it probably is strain specific.

I designed an experiment to test for feedback activation of Lrp in an additional strain of *V. cholerae* (C6706) to determine if there is conservation of an activation phenotype in Vibrionaceae family. I carried out a $\beta$-galactosidase assay to test the expression of the *V. cholerae* P*lrp* promoter in *lrp*+ and *lrp*− backgrounds. Expression levels in all Vibrio strains tested were very low and it is not certain how much of an effect was produced by the P*lrp*-LacZ promoter fusion. The effect produced is hard to critically examine and this is as a result of low baseline seen in Fig. 1. In this study, all strains (*lrp* and *lrp*+) produced a repression phenotype in the presence or absence of Lrp (Fig 1). Transcription levels were very low as can be seen in the data (Fig 1). The repression phenotype produced by the strains is not consistent with the activation phenotype seen previously [14].

Interestingly, our results differ from the trend seen in previous studies, and this might be attributable to the use of LB medium as opposed to the previous use of minimal medium. The choice of medium has been shown to have impact on transcription level in bacteria. LB medium has been a selective choice for bacteria growth but gives varying effect in studying transcription mechanism. In *E. coli*, two major factors affect
transcription: growth rate and associated changes in guanosine 3', 5'-bispyrophosphate (ppGpp) levels [36, 37], and feedback repression by Lrp on its own gene [32]. Previous studies by Lintner et. al used MOPS glucose medium to test the regulation of Plrp while the current study used LB medium. Rich medium such as LB contains an undefined mix of amino acids, vitamins, and other nutrients, and in E. coli results in depressed levels of ppGpp [37]. In addition, and more specifically, LB contains the amino acids leucine, alanine, and methionine, each of which have major effects on Lrp activity [31]. It is likely that the components of LB rich medium affected the expression of lrp. Though Lrp repression of the lrp gene itself is relatively insensitive to coregulatory leucine when added to minimal medium; growth in LB greatly reduces lrp transcription [32, 38]. In other words, Lrp can autorepress independently of a coregulator, via its effect on growth rate via ppGpp. In LB medium, lrp transcription is expected to be low since ppGpp stimulates transcription of lrp, and ppGpp levels are lowest in rich medium. The strong Plrp-lacZ expression in the V. cholerae C6706 lrp background, in LB, is a surprise, as is the fact that Lrp represses in this particular background. Comparing the present results in two vibrio strains to previous findings, the choice of medium is a variable that needs to be tested in the future. The use of LB medium affects the given transcription level due to its components.

Another possible explanation for the difference between the two studies is the strain used. Lintner et al. used strain A1552, which has been used in a number of studies [39], while here the more commonly-used strain C6706 was used [40]. It is unclear how the strains differ, as both are of the El Tor Inaba biotype [41, 42]. While a genome sequence of strain A1552 is not currently available, it is unlikely that sequence
differences in the Plrp region explain the different behavior observed here, as that region is generally well conserved among the Vibrionaceae (Fig. 2). We believe that regulatory effects of Lrp is strain specific in *V. cholerae*.

In future studies, we will use minimal media to repeat this experiment, to eliminate any form of bias and keep all experimental conditions constant as it has been earlier stated that LB medium affects transcription. Our results present findings already known about regulation of Lrp but also points out the effects of using LB rich medium. This study has been able to confirm findings on the regulation of Lrp in *Vibrio cholerae*. 
2.6 References


Figure 2.1: Lrp expression in Vibrio cholerae. Plasmid were transformed into the lrp<sup>+</sup> strains and LacZ activity was measured in Miller units. Numbers above the error bars are calculated miller units of each strain. (N. S – not significant)
**Figure 2.2. Conservation of Plrp region in Vibrionaceae.** Sequence alignment of lrp promoter regions in different Vibrionaceae strains. Two regions highlighted in yellow and green correspond to Shine-Dalgarno ribosomal binding sequence and initiation codon for the Lrp Open reading frame.
Chapter 3

Conclusion

This thesis presents data from our studies on the global regulator Lrp. The first study explores the evolution of Lrp in four related bacterial orders based on phylogenetic and logo analyses. Lrp plays alternative roles (local or global) in different bacterial orders. Our goal in studying the evolution of Lrp using these analyses is that they provide a good foundation for studying the basis for global vs. local roles of Lrp and other regulators in the future. Our results highlighted the relationship between Lrp and its paralog AsnC. We found that Lrp shows a similar phylogenetic relationship in three bacterial orders, which closely follows that of its host core genome based on two representative housekeeping genes. However, one of the four bacterial orders analyzed, Alteromonadales, was unique in showing variation in its phylogenetic clustering pattern for \textit{lrp}, \textit{asnC}, and the housekeeping gene \textit{recA}, even though 16S rRNA and \textit{rpoB} clustered as expected. The Alteromonadales are a diverse group with poorly characterized taxonomic representatives. The different clustering patterns observed in our analysis of the Alteromonadales could reflect differences in ecological niches or an unusually rapid rate at which horizontal gene transfer might occur in this order. Our analysis of sequence signatures associated with globally- vs. locally-acting Lrp orthologs (see below) is
consistent with the possibility that different Alteromonadales *lrp* clusters play different roles (global vs. local). A detailed genomic analysis of additional species from the Alteromonadales, along with assessment of how many genes Lrp controls in each, would give us more insight.

Another part of our study focuses on logo analysis of Lrp in four bacterial orders. Identifying possible sequence motifs that could readily distinguish Lrp acting as global or local regulator could provide more insight into protein function on a sequence level. This analysis also allowed us to identify cases of gene misannotation between Lrp and AsnC. Due to the automated process of annotation for databases, this process is error prone. Lrp and AsnC are members of the same family and likely candidates for gene annotation error as we routinely observed in our studies. Therefore, we suggest that Logo analysis can serve as a good technique for identifying and correcting gene misannotations.

Conserved sequence motifs were identified that may serve as targets for future investigation. Some important residue positions identified are areas involved in possible post transcriptional modification such as lysine acetylation and coregulator binding. Other identified motifs with long stretches of amino acid residues have unknown functions, and we believe that these conserved motifs that are either unique to a specific bacterial order or to a functional role are important in understanding Lrp activity in species where the role of Lrp has not been tested. Functional analysis of these identified conserved motifs could serve as a missing piece to the puzzle of alternating roles of Lrp.

We also performed studies on Lrp regulation of the *lrp* promoter (autogenous control) in different *Vibrio cholerae* strains. We carried out this study based on a previous finding that *lrp* orthologs in *E. coli* and *V. cholerae* regulate their own genes differently.
Lrp represses expression of its own gene in *E. coli* while it activates its expression in one tested strain of *V. cholerae*. Using these previous findings, I was interested in determining if this surprising phenotype is conserved in the Vibrionaceae family. Upon examining the regulation of *lrp* expression by Lrp in an additional strain of *V. cholerae*, we found that the two strains did not show the same regulation pattern. The other tested strain shows a similar regulatory pattern to that of *E. coli*. Therefore, we conclude that the regulation of the *lrp* promoter by Lrp in Vibrionaceae might be strain dependent. This indicates that not only the intrinsic activity of Lrp, but also its amounts in the cell, likely show substantial variation between species. For a transcription factor that can control hundreds of genes, this is a significant observation.
References


(b) Chapter 2


