

Functional Diversity in the PIP Subfamily: Insights from Sequence Similarity Networks

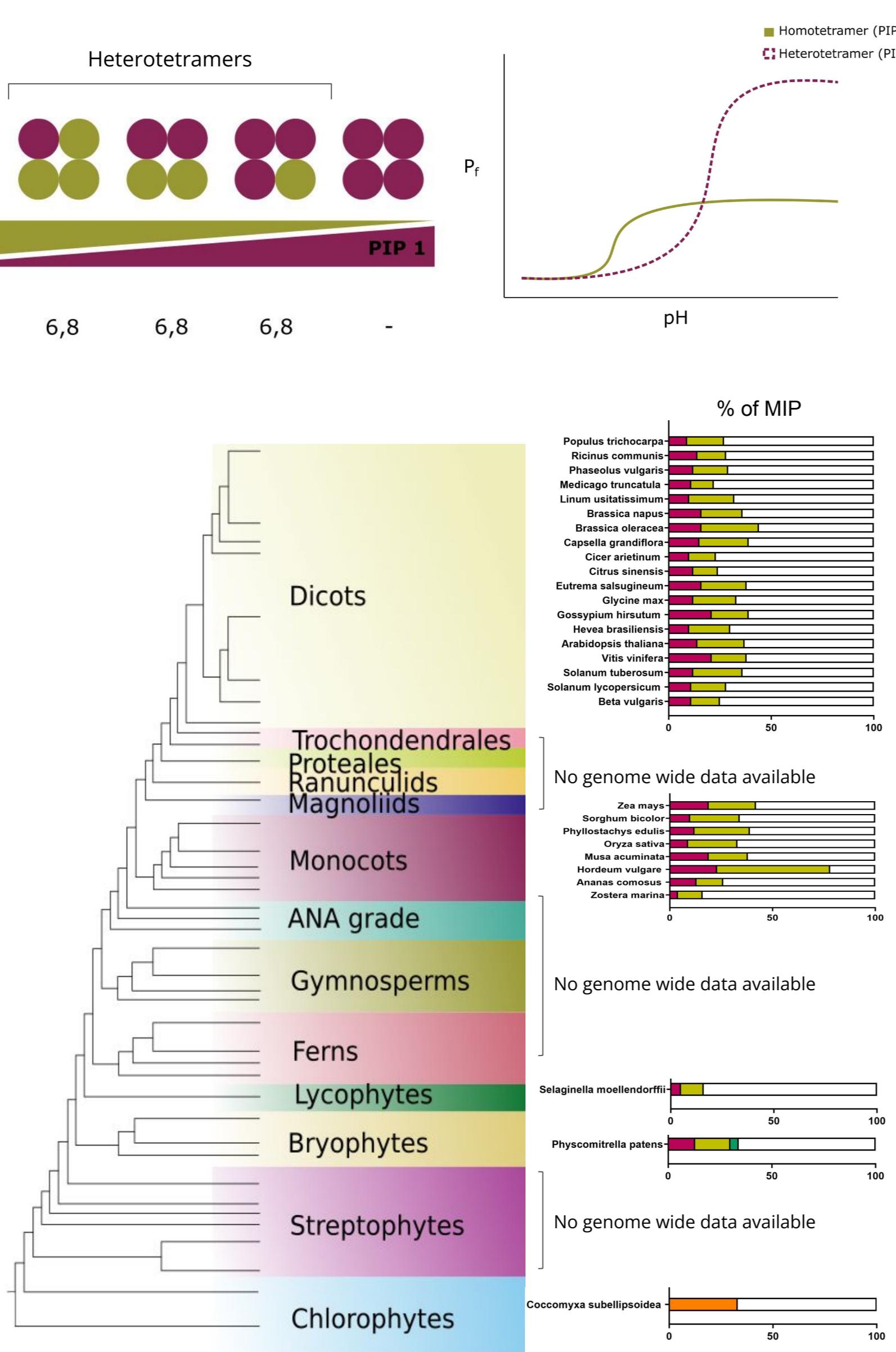
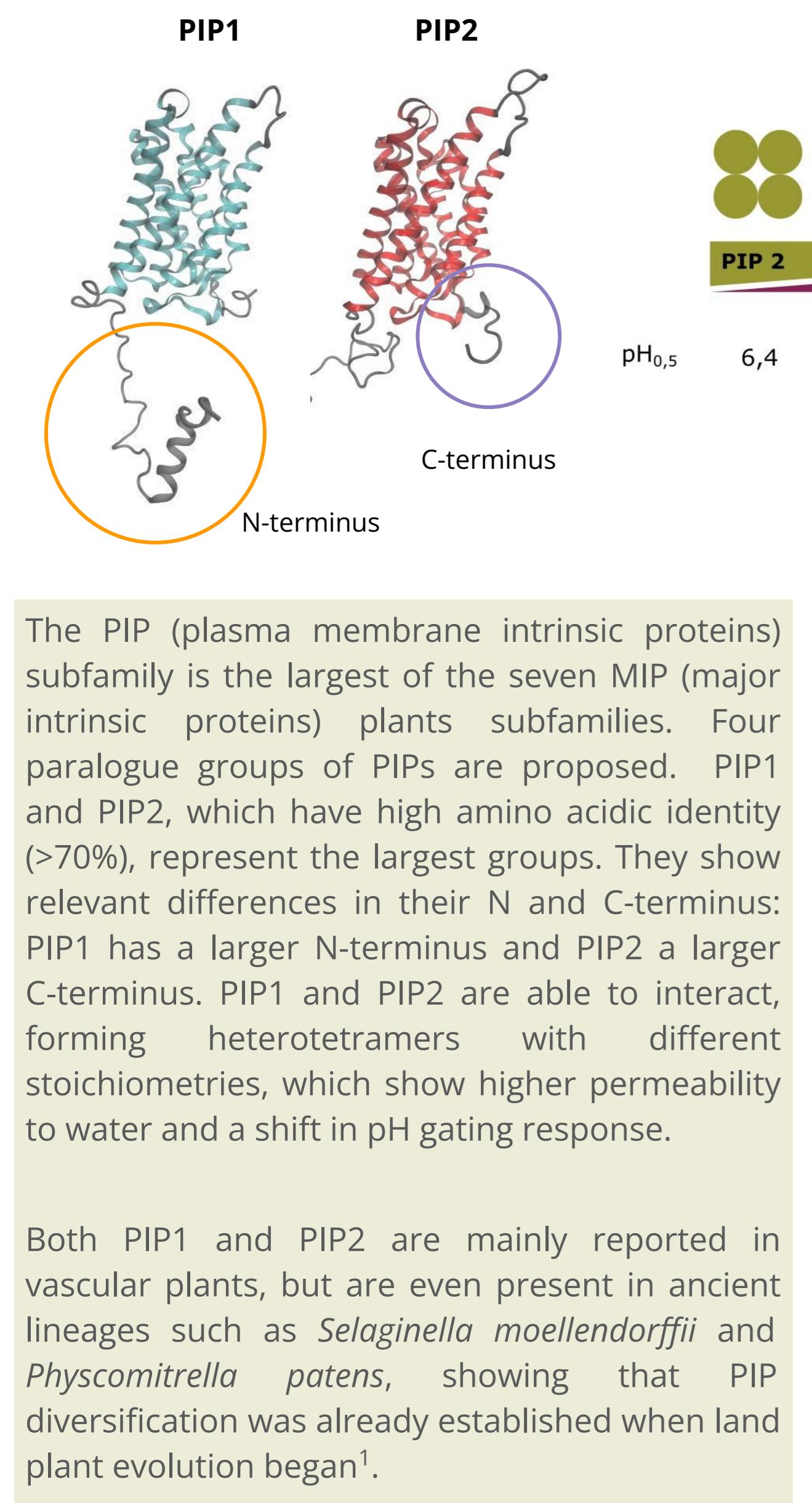
Bussolini Lizundia R¹, Fox R², Vitali V^{1,3}, Alvea K^{1,3}.

¹Universidad de Buenos Aires, Facultad de Farmacia y Bioquímica, Departamento de Fisicomatemática, Cátedra de Física, Buenos Aires, Argentina. ²Institut des Sciences de la Vie, Université catholique de Louvain, Croix du Sud 4-L7.07.14, B-1348 Louvain-la-Neuve, Belgium ³Universidad de Buenos Aires-CONICET, Instituto de Química y Fisicoquímica Biológicas "Prof. Alejandro C. Paladini", Buenos Aires, Argentina.

rbussolini@docente.ffy.uba.ar

PICT-2019-00387

1. PIP Subfamily

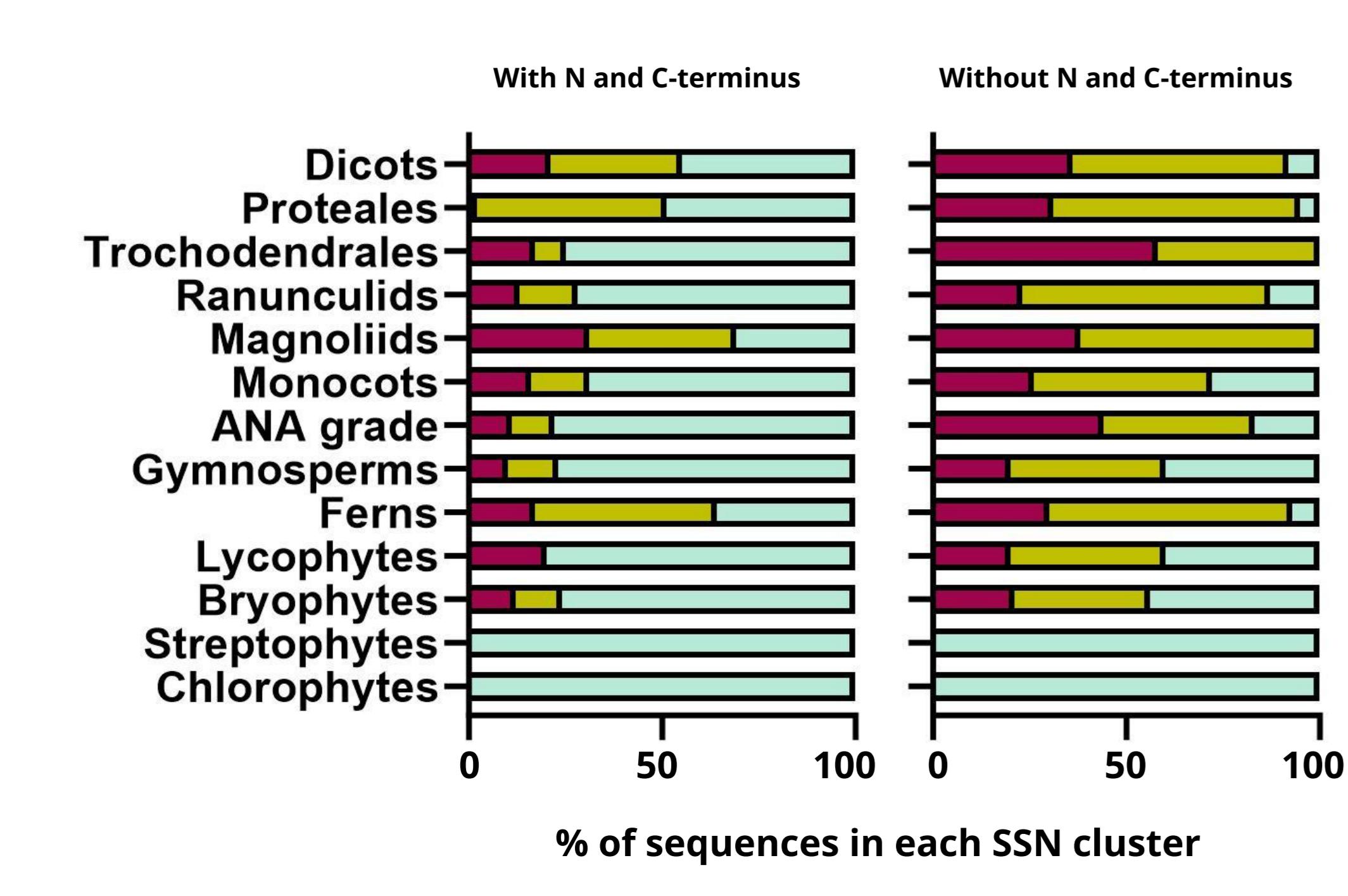
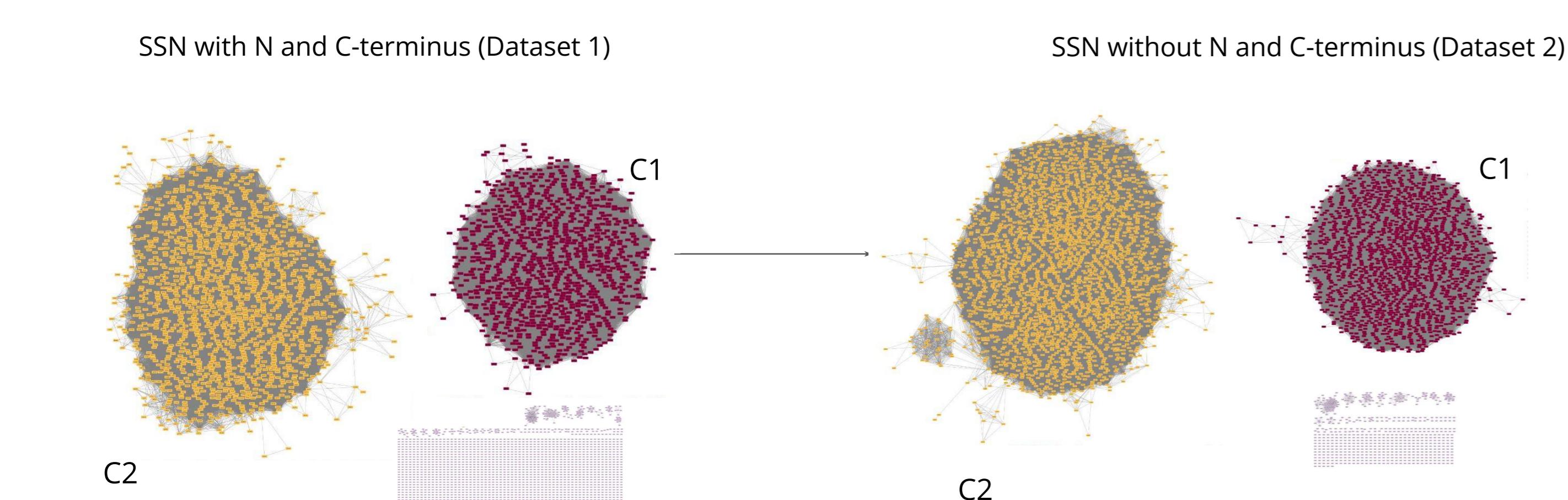


3. Do the N and C-terminus play a role in PIP1 and PIP2 differential clustering?

Based on the relevance of the N and C-termini between PIP paralogous, we created another dataset (Dataset 2) in which all sequences had their N and C-termini removed. With N and C-termini removal, 8.85% of the dataset's sequences became redundant, indicating that some PIP sequences only differed in those regions.



Comparing the SSNs obtained for each dataset, upon N and C-terminus removal PIP1 and PIP2 still clusterize separately, forming two big clusters. This means that the N and C-terminus are not the only motifs that determine PIP isoforms differential clustering. Their deletion also enriched C1 and C2 with sequences that had been miscellaneous in dataset 1.

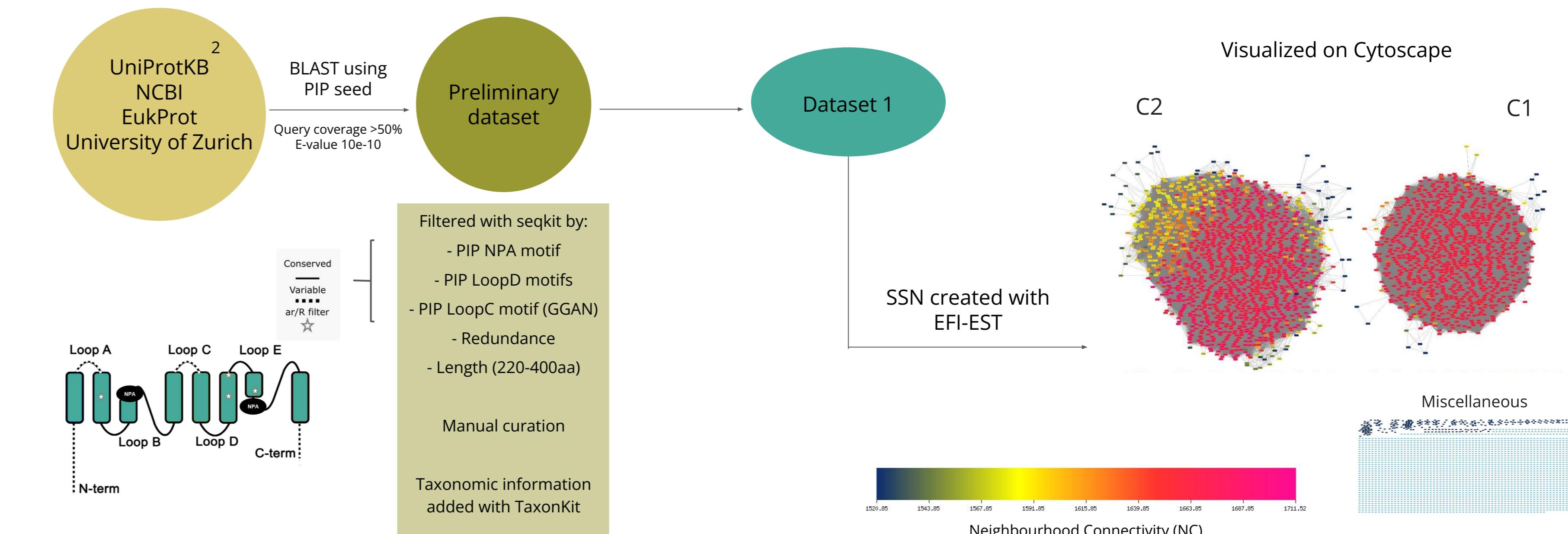


Conclusions and perspectives

- SSNs facilitated the analysis of the PIP subfamily, as it simplifies working with large datasets and studying them from multiple perspectives.
- It exhibited how the N and C-terminus motifs, which are often regarded as the main differences within PIP1 and PIP2, are not the only motifs that differentiate both isoform clusters. The SSN also exposed that PIP2s and PIP1s may have different constraints during evolution. The modification of the NC threshold divided the PIP2 cluster (C2) in three clusters without a major effect in the PIP1 cluster (C1).
- Analyzing differences and similarities between clusters is important to understand diverse functional properties of PIP isoform clusters. Clusters of the PIP subfamily proved to be multifunctional, leading to consider that the multiple sequences within them with no experimental data or transport reported for just one solute may have functions that have yet to be proved.
- Future analysis will focus on specifically characterizing what motifs or residues lead to PIP differential clusterization. PIP representation and distribution among different lineages will also be explored, especially for ancient lineages which have not been studied at the same extent land plants have.

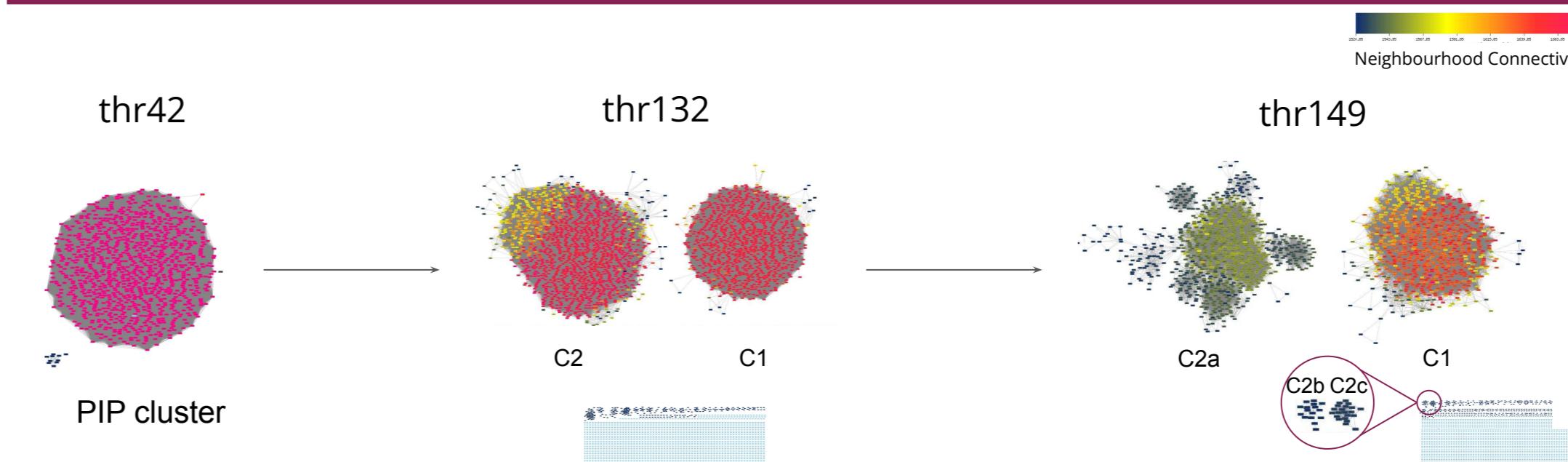
2. How to identify PIP amino acid sequences?

PIP subfamily lacks a specific tag to find its sequences throughout databases, so a different approach was needed to identify them. A set of well characterized PIPs (PIP seed) were utilized as query to blast in different databases and create a preliminary PIP dataset. This dataset was scanned and filtered with signature PIP motifs to get the PIP subfamily subset using the seqkit software. These included the NPA, LoopC and LoopD (involved in gating mechanism) motifs. For the LoopD motif, a consensus sequence by manual curation was used, as well as the altered LoopD of a characterized *Coccomyxa subellipsoidea* PIP isoform. The database was enriched with taxonomic information using TaxonKit.

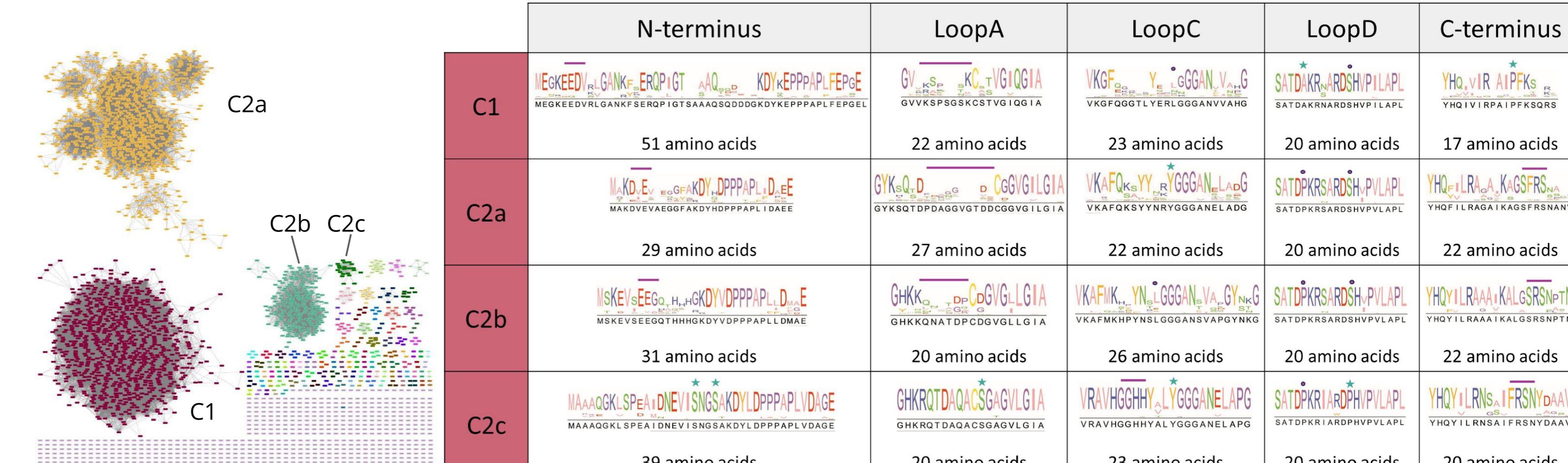


A sequence similarity network (SSN) was built using the EFI-EST³ online tool and visualized in Cytoscape v3.10.0. In the network showed each amino acid sequence is represented by a node. PIP subfamily topology in the SSN was characterized, with PIP2 isoforms organizing into multiple clusters, including a significant neighbourhood connectivity (NC) heterogeneity in the largest cluster (C2). On the other hand, PIP1 representatives formed one large, homogeneous cluster (C1) with high NC.

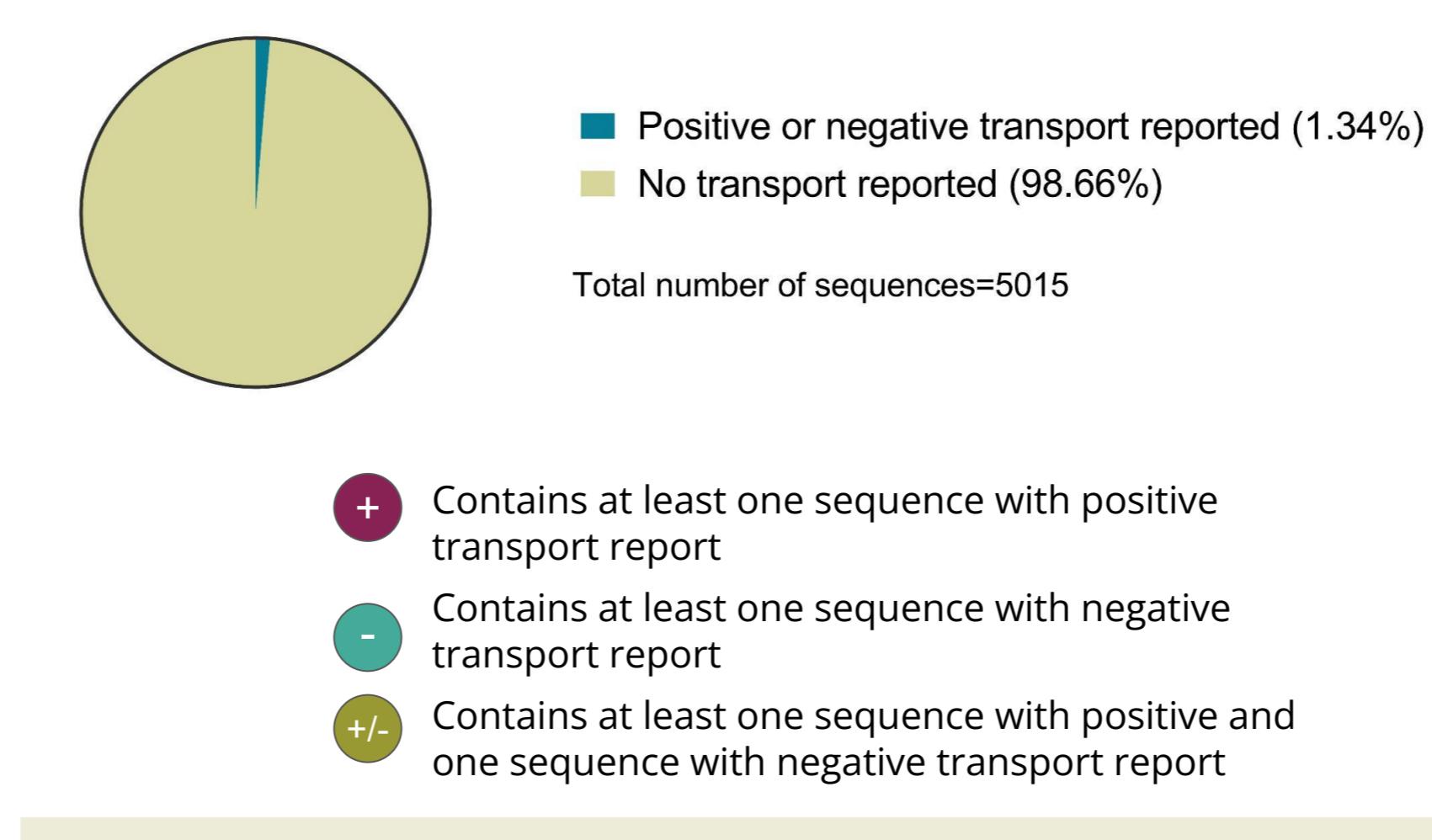
4. Sequence diversity analysis: What characterizes each cluster?



C1 and C2a were analysed as well as C2b and C2c, two clusters derived from C2 in thr132 in which two well characterized PIPs from the PIP seed clustered. Variable N and C-terminus, LoopA and LoopC were studied along the highly conserved LoopD to evaluate their differences between each of these clusters. Amino acid sequence logos were obtained for each of these motifs.



N-terminus exhibited different lengths and regions of high variability, which can be also related to insertions. In C2c, the diacidic motif⁴ observed was absent and it presented two highly conserved serines. LoopA obtained for the other three clusters showed its high variability, and how C2b and C2c differed in length from C2a. The GGAN LoopC motif was preceded by a tyrosine in both C2a and C2c, while the latter also presented a conserved GGHH motif in this loop which did not appear in other clusters. LoopD is highly conserved amongst all four clusters, with some varying residues as shown in the image. C1 had a shorter C-terminus than the other clusters, with a conserved proline which C2a, C2b and C2c did not have. These three clusters showed a PIP2 (S/F)RSN motif.



Their distribution within the network suggests that the clusters obtained are multifunctional, with PIPs with positive and/or negative transport of different solutes in each of them.

To see whether these clusters and their motifs were related to a particular solute transport, which would make them isofunctional clusters, we analysed positive and negative transport reported for PIPs, which comprehends only 1,34% of the dataset.

	H ₂ O	H ₂ O ₂	Glycerol	CO ₂	Na ⁺	Boron	O ₂
C1	+/-	+/-		+		+/-	
C2a	+	+/-	+	+/-	+/-	-	
C2b	+/-						
C2c	+					+	
Other clusters/ singletons	+/-	+/-	+/-	+	+/-	+/-	+/-

- References:
1. Anderberg HJ, Kjellbom P, Johanson U. Annotation of *Selaginella moellendorffii* Major Intrinsic Proteins and the Evolution of the Protein Family in Terrestrial Plants. *Front Plant Sci*. 2012 Feb 20;3:33. doi: 10.3389/fpls.2012.00033. PMID: 22639644; PMCID: PMC335642.
 2. <https://www.uniprot.org/uniprotkb>
 3. <https://www.ncbi.nlm.nih.gov/>
 4. <https://www.eurocellbio.com/eukprot/>
 5. <https://www.hornworts.uzh.ch/en/download.html>
 6. <https://www.ncbi.nlm.nih.gov/>
 7. <https://www.ncbi.nlm.nih.gov/>
 8. <https://www.ncbi.nlm.nih.gov/>
 9. <https://www.ncbi.nlm.nih.gov/>
 10. <https://www.ncbi.nlm.nih.gov/>
 11. <https://www.ncbi.nlm.nih.gov/>
 12. <https://www.ncbi.nlm.nih.gov/>
 13. <https://www.ncbi.nlm.nih.gov/>
 14. <https://www.ncbi.nlm.nih.gov/>
 15. <https://www.ncbi.nlm.nih.gov/>
 16. <https://www.ncbi.nlm.nih.gov/>
 17. <https://www.ncbi.nlm.nih.gov/>
 18. <https://www.ncbi.nlm.nih.gov/>
 19. <https://www.ncbi.nlm.nih.gov/>
 20. <https://www.ncbi.nlm.nih.gov/>
 21. <https://www.ncbi.nlm.nih.gov/>
 22. <https://www.ncbi.nlm.nih.gov/>
 23. <https://www.ncbi.nlm.nih.gov/>
 24. <https://www.ncbi.nlm.nih.gov/>
 25. <https://www.ncbi.nlm.nih.gov/>
 26. <https://www.ncbi.nlm.nih.gov/>
 27. <https://www.ncbi.nlm.nih.gov/>
 28. <https://www.ncbi.nlm.nih.gov/>
 29. <https://www.ncbi.nlm.nih.gov/>
 30. <https://www.ncbi.nlm.nih.gov/>
 31. <https://www.ncbi.nlm.nih.gov/>
 32. <https://www.ncbi.nlm.nih.gov/>
 33. <https://www.ncbi.nlm.nih.gov/>
 34. <https://www.ncbi.nlm.nih.gov/>
 35. <https://www.ncbi.nlm.nih.gov/>
 36. <https://www.ncbi.nlm.nih.gov/>
 37. <https://www.ncbi.nlm.nih.gov/>
 38. <https://www.ncbi.nlm.nih.gov/>
 39. <https://www.ncbi.nlm.nih.gov/>
 40. <https://www.ncbi.nlm.nih.gov/>
 41. <https://www.ncbi.nlm.nih.gov/>
 42. <https://www.ncbi.nlm.nih.gov/>
 43. <https://www.ncbi.nlm.nih.gov/>
 44. <https://www.ncbi.nlm.nih.gov/>
 45. <https://www.ncbi.nlm.nih.gov/>
 46. <https://www.ncbi.nlm.nih.gov/>
 47. <https://www.ncbi.nlm.nih.gov/>
 48. <https://www.ncbi.nlm.nih.gov/>
 49. <https://www.ncbi.nlm.nih.gov/>
 50. <https://www.ncbi.nlm.nih.gov/>
 51. <https://www.ncbi.nlm.nih.gov/>
 52. <https://www.ncbi.nlm.nih.gov/>
 53. <https://www.ncbi.nlm.nih.gov/>
 54. <https://www.ncbi.nlm.nih.gov/>
 55. <https://www.ncbi.nlm.nih.gov/>
 56. <https://www.ncbi.nlm.nih.gov/>
 57. <https://www.ncbi.nlm.nih.gov/>
 58. <https://www.ncbi.nlm.nih.gov/>
 59. <https://www.ncbi.nlm.nih.gov/>
 60. <https://www.ncbi.nlm.nih.gov/>
 61. <https://www.ncbi.nlm.nih.gov/>
 62. <https://www.ncbi.nlm.nih.gov/>
 63. <https://www.ncbi.nlm.nih.gov/>
 64. <https://www.ncbi.nlm.nih.gov/>
 65. <https://www.ncbi.nlm.nih.gov/>
 66. <https://www.ncbi.nlm.nih.gov/>
 67. <https://www.ncbi.nlm.nih.gov/>
 68. <https://www.ncbi.nlm.nih.gov/>
 69. <https://www.ncbi.nlm.nih.gov/>
 70. <https://www.ncbi.nlm.nih.gov/>
 71. <https://www.ncbi.nlm.nih.gov/>
 72. <https://www.ncbi.nlm.nih.gov/>
 73. <https://www.ncbi.nlm.nih.gov/>
 74. <https://www.ncbi.nlm.nih.gov/>
 75. <https://www.ncbi.nlm.nih.gov/>
 76. <https://www.ncbi.nlm.nih.gov/>
 77. <https://www.ncbi.nlm.nih.gov/>
 78. <https://www.ncbi.nlm.nih.gov/>
 79. <https://www.ncbi.nlm.nih.gov/>
 80. <https://www.ncbi.nlm.nih.gov/>
 81. <https://www.ncbi.nlm.nih.gov/>
 82. <https://www.ncbi.nlm.nih.gov/>
 83. <https://www.ncbi.nlm.nih.gov/>
 84. <https://www.ncbi.nlm.nih.gov/>
 85. <https://www.ncbi.nlm.nih.gov/>
 86. <https://www.ncbi.nlm.nih.gov/>
 87. <https://www.ncbi.nlm.nih.gov/>
 88. <https://www.ncbi.nlm.nih.gov/>
 89. <https://www.ncbi.nlm.nih.gov/>
 90. <https://www.ncbi.nlm.nih.gov/>
 91. <https://www.ncbi.nlm.nih.gov/>
 92. <https://www.ncbi.nlm.nih.gov/>
 93. <https://www.ncbi.nlm.nih.gov/>
 94. <https://www.ncbi.nlm.nih.gov/>
 95. <https://www.ncbi.nlm.nih.gov/>
 96. <https://www.ncbi.nlm.nih.gov/>
 97. <https://www.ncbi.nlm.nih.gov/>
 98. <https://www.ncbi.nlm.nih.gov/>
 99. <https://www.ncbi.nlm.nih.gov/>
 100. <https://www.ncbi.nlm.nih.gov/>
 101. <https://www.ncbi.nlm.nih.gov/>
 102. <https://www.ncbi.nlm.nih.gov/>
 103. <https://www.ncbi.nlm.nih.gov/>
 104. <https://www.ncbi.nlm.nih.gov/>
 105. <https://www.ncbi.nlm.nih.gov/>
 106. <https://www.ncbi.nlm.nih.gov/>
 107. <https://www.ncbi.nlm.nih.gov/>
 108. <https://www.ncbi.nlm.nih.gov/>
 109. <https://www.ncbi.nlm.nih.gov/>
<li