Parent-map: analysis of parental contributions to evolved or engineered protein or DNA sequences

Damien Marsic[1]

1 Porton Biologics, 388 Xinping Street, Suzhou Industrial Park, Jiangsu 215021, China

Summary

Parent-map analyzes protein or DNA sequences which are derived from one or multiple parent sequences, and shows parental contributions as well as differences from relevant parents. Originally developed to analyze capsid protein sequences obtained by directed evolution, parent-map can be used in any case where variant sequences are to be compared to parent sequences from which they are derived. Parent-map detects sequence shuffling as well as substitutions, insertions and deletions, and displays results in user-friendly formats. Parent-map is an open-source, platform-independent Python 3 script, available as a Bioconda package as well as a Windows program.

Source code: https://github.com/damienmarsic/Parent-map

Python package: https://pypi.org/project/parent-map/

Bioconda recipe and package: http://bioconda.github.io/recipes/parent-map/README.html

Windows installer: https://sourceforge.net/projects/parent-map/

Documentation: https://parent-map.readthedocs.io/

Statement of need

Adeno-associated virus (AAV) capsid directed evolution projects typically generate multiple enriched variant sequences after 2 to 5 rounds of selection starting from complex capsid libraries. For libraries developed from a single parental serotype, through random peptide insertion at a specific position or surface loop diversification in well-defined variable regions for example, a single multiple alignment of all enriched variant sequences against the parent sequence conveniently shows how each variant differs from the parent. However, when more than one parental sequence is involved, such as when different libraries are mixed together, or when a library design involves DNA shuffling from several parents, such alignments can quickly become illegible, particularly when the complete capsid gene is sequenced. In such cases, in the absence of appropriate software tools, each variant needs to be separately aligned against all possible parents, a time-consuming and cumbersome process. An added difficulty in the case of shuffled libraries is that, because of high sequence homology between parents, multiple regions will share sequence identities with more than one parent, complicating attempts at comprehensively defining the variant sequences in terms of parental contributions. To date, SALANTO (Herrmann et al., 2019) seems to be the only relevant publicly available software. However, it only applies to shuffled libraries, and its user-friendliness is limited as it requires the user to perform a multiple sequence alignment beforehand, and to further process the data manually after analysis. The software described in this article, parent-map, provides a user-friendly and comprehensive solution. It can be used with sequences derived from any

type of library, or even with naturally-occurring mutants or rationally engineered variants. It is not limited to protein sequences. It only requires one file containing the variant sequences to be analyzed, and one file containing parental sequences, without any prior manipulation. It generates a set of five files covering most end-users' needs, in directly usable formats. Finally, although it was developed to address a need in the field of AAV capsid directed evolution, parent-map can be used whenever protein or DNA sequences, whether originating from natural evolution, directed evolution or rational design, are to be compared with one or more possible parental sequences.

Methods

Parent-map was written under Python 3.7 as both a command-line interface (CLI) and a graphical user interface (GUI) application, by allowing parser modules argparse and Gooey to coexist within a single file (the GUI will start if no argument is present, while any argument will cause parent-map to start in CLI mode). A parent-map Python package was created and uploaded to the Python Package Index (PyPI) according to packaging instructions. A parent-map Bioconda (Grüning et al., 2018) recipe based on the PyPI package was written and submitted according to instructions. A stand-alone Windows executable and its installation program were created using respectively PyInstaller and Inno Setup. The documentation was written using Sphinx.

Implementation

Parent-map is a platform-independent Python script that generates a set of five output files from two input files. Input file names and options can be entered as arguments at launch time, resulting in parent-map running in CLI mode, or within the GUI, which starts if parent-map is launched without arguments. This flexibility allows parent-map to be deployed in a variety of settings, as a simple desktop application or even as a bioinformatics pipeline component. The first input file contains the variant sequences, typically the most frequent or the most enriched sequences obtained at the completion of a directed evolution experiment. The other input file is a set of potential parental sequences to the variant sequences. The most useful files generated by parent-map, particularly in the case of variants derived from DNA shuffling, are parental contribution maps (file names ending in –par.txt and –par.html, the latter being a colorized version of the former). Instead of all possible combinations, the simplest map that can accurately describe the variant is shown, using as few parents and as few fragments as possible. Other output files include a statistics file summarizing the variant sequences main features, a sequence definition file comprehensively defining each variant in terms of its parents, and an alignment file showing how variants differ from their common parent.

Parent-map can be tested using the provided variant and parent sample files, based on available literature describing evolved and rationally designed AAV capsid variants. Variants AAV-DJ (Grimm et al., 2008), AAV2.5T (Excoffon et al., 2009), NP84 (Paulk et al., 2018) and OLIG001 (Powell et al., 2016) are derived from shuffled DNA libraries. Variants AAV-F (Hanlon et al., 2019), AAV-PHP.B (Deverman et al., 2016), 7m8 (Dalkara et al., 2013) and rAAV2-retro (Tervo et al., 2016) are derived from peptide insertion libraries. Variants SCH2, SCH9 (Ojala et al., 2018), LI-A and LI-C (Marsic et al., 2014) are derived from more complex rationally designed libraries. Variants AAV2i8 (Asokan et al., 2010) and AAV2-sept-Y-F (Petrs-Silva et al., 2011) were rationally designed. Using default settings, parent-map correctly identifies single parental contributions from AAV9 for variants AAV-F and AAV-PHP.B, single parental contributions from AAV2 for variants 7m8, rAAV2-retro, LI-A, LI-C, AAV2-sept-Y-F, and multiple parental contributions from AAV2, AAV8 and AAV9 for AAV-DJ, from AAV2 and AAV5 for AAV2.5T, from AAV2, AAV3B and AAV6 for NP84, from

AAV2, AAV6, AAV8 and AAV9 for OLIG001, SCH2 and SCH9, and from AAV2 and AAV8 for AAV2i8. Parent-map also correctly detects peptide insertions FVVGQSY for AAV-F and TLAVPFK for AAV-PHP.B, both at position 588, and peptide insertions LALGETTRPA for 7m8 and LADQDYTKTA for rAAV2-retro, both at position 587. Finally, parent-map correctly identifies substitutions A to T at position 457 for AAV-DJ and at position 582 for AAV2.5T, substitutions K to E at 532 and R to G at 585 for NP84, E to K substitution at 532 and unmatched H at 726 for OLIG001, substitutions I to T at 240 and V to I at 718 for 7m8, substitutions N to D at 382 and V to I at 718 for rAAV2-retro, the 14 and 4 substitutions for LI-A and LI-C respectively, as well as the 7 Y to F substitutions at 252, 272, 444, 500, 700, 704 and 730 for AAV2-sept-Y-F.

A comprehensive description of parent-map is provided in the documentation.

Acknowledgements

References

Asokan, A., Conway, J. C., Phillips, J. L., Li, C., Hegge, J., Sinnott, R., Yadav, S., DiPrimio, N., Nam, H.-J., Agbandje-McKenna, M., McPhee, S., Wolff, J., & Samulski, R. J. (2010). Reengineering a receptor footprint of adeno-associated virus enables selective and systemic gene transfer to muscle. *Nature Biotechnology*, *28*(1), 79–82. https://doi.org/10.1038/nbt.1599

Dalkara, D., Byrne, L. C., Klimczak, R. R., Visel, M., Yin, L., Merigan, W. H., Flannery, J. G., & Schaffer, D. V. (2013). In Vivo-Directed Evolution of a New Adeno-Associated Virus for Therapeutic Outer Retinal Gene Delivery from the Vitreous. *Science Translational Medicine*, *5*(189), 189ra76–189ra76. https://doi.org/10.1126/scitranslmed.3005708

Deverman, B. E., Pravdo, P. L., Simpson, B. P., Kumar, S. R., Chan, K. Y., Banerjee, A., Wu, W.-L., Yang, B., Huber, N., Pasca, S. P., & Gradinaru, V. (2016). Cre-dependent selection yields AAV variants for widespread gene transfer to the adult brain. *Nature Biotechnology*, *34*(2), 204–209. https://doi.org/10.1038/nbt.3440

Excoffon, K. J. D. A., Koerber, J. T., Dickey, D. D., Murtha, M., Keshavjee, S., Kaspar, B. K., Zabner, J., & Schaffer, D. V. (2009). Directed evolution of adeno-associated virus to an infectious respiratory virus. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(10), 3865–3870. https://doi.org/10.1073/pnas.0813365106

Grimm, D., Lee, J. S., Wang, L., Desai, T., Akache, B., Storm, T. A., & Kay, M. A. (2008). In Vitro and In Vivo Gene Therapy Vector Evolution via Multispecies Interbreeding and Retargeting of Adeno-Associated Viruses. *Journal of Virology*, *82*(12), 5887–5911. https://doi.org/10.1128/JVI.00254-08

Grüning, B., Dale, R., Sjödin, A., Chapman, B. A., Rowe, J., Tomkins-Tinch, C. H., Valieris, R., Köster, J., & Bioconda Team. (2018). Bioconda: Sustainable and comprehensive software distribution for the life sciences. *Nature Methods*, *15*(7), 475–476. https://doi.org/10.1038/s41592-018-0046-7

Hanlon, K. S., Meltzer, J. C., Buzhdygan, T., Cheng, M. J., Sena-Esteves, M., Bennett, R. E., Sullivan, T. P., Razmpour, R., Gong, Y., Ng, C., Nammour, J., Maiz, D., Dujardin, S., Ramirez, S. H., Hudry, E., & Maguire, C. A. (2019). Selection of an Efficient AAV

Vector for Robust CNS Transgene Expression. *Molecular Therapy. Methods & Clinical Development*, *15*, 320–332. https://doi.org/10.1016/j.omtm.2019.10.007

Herrmann, A.-K., Bender, C., Kienle, E., Grosse, S., El Andari, J., Botta, J., Schürmann, N., Wiedtke, E., Niopek, D., & Grimm, D. (2019). A Robust and All-Inclusive Pipeline for Shuffling of Adeno-Associated Viruses. *ACS Synthetic Biology*, *8*(1), 194–206. https://doi.org/10.1021/acssynbio.8b00373

Marsic, D., Govindasamy, L., Currlin, S., Markusic, D. M., Tseng, Y.-S., Herzog, R. W., Agbandje-McKenna, M., & Zolotukhin, S. (2014). Vector design Tour de Force: Integrating combinatorial and rational approaches to derive novel adeno-associated virus variants. *Molecular Therapy: The Journal of the American Society of Gene Therapy*, *22*(11), 1900–1909. https://doi.org/10.1038/mt.2014.139

Ojala, D. S., Sun, S., Santiago-Ortiz, J. L., Shapiro, M. G., Romero, P. A., & Schaffer, D. V. (2018). In Vivo Selection of a Computationally Designed SCHEMA AAV Library Yields a Novel Variant for Infection of Adult Neural Stem Cells in the SVZ. *Molecular Therapy: The Journal of the American Society of Gene Therapy*, *26*(1), 304–319. https://doi.org/10.1016/j.ymthe.2017.09.006

Paulk, N. K., Pekrun, K., Zhu, E., Nygaard, S., Li, B., Xu, J., Chu, K., Leborgne, C., Dane, A. P., Haft, A., Zhang, Y., Zhang, F., Morton, C., Valentine, M. B., Davidoff, A. M., Nathwani, A. C., Mingozzi, F., Grompe, M., Alexander, I. E., … Kay, M. A. (2018). Bioengineered AAV Capsids with Combined High Human Liver Transduction In Vivo and Unique Humoral Seroreactivity. *Molecular Therapy: The Journal of the American Society of Gene Therapy*, *26*(1), 289–303. https://doi.org/10.1016/j.ymthe.2017.09.021

Petrs-Silva, H., Dinculescu, A., Li, Q., Deng, W.-T., Pang, J.-J., Min, S.-H., Chiodo, V., Neeley, A. W., Govindasamy, L., Bennett, A., Agbandje-McKenna, M., Zhong, L., Li, B., Jayandharan, G. R., Srivastava, A., Lewin, A. S., & Hauswirth, W. W. (2011). Novel properties of tyrosine-mutant AAV2 vectors in the mouse retina. *Molecular Therapy: The Journal of the American Society of Gene Therapy*, *19*(2), 293–301. https://doi.org/10.1038/mt.2010.234

Powell, S. K., Khan, N., Parker, C. L., Samulski, R. J., Matsushima, G., Gray, S. J., & McCown, T. J. (2016). Characterization of a novel adeno-associated viral vector with preferential oligodendrocyte tropism. *Gene Therapy*, *23*(11), 807–814. https://doi.org/10.1038/gt.2016.62

Tervo, D. G. R., Hwang, B.-Y., Viswanathan, S., Gaj, T., Lavzin, M., Ritola, K. D., Lindo, S., Michael, S., Kuleshova, E., Ojala, D., Huang, C.-C., Gerfen, C. R., Schiller, J., Dudman, J. T., Hantman, A. W., Looger, L. L., Schaffer, D. V., & Karpova, A. Y. (2016). A Designer AAV Variant Permits Efficient Retrograde Access to Projection Neurons. *Neuron*, *92*(2), 372–382. https://doi.org/10.1016/j.neuron.2016.09.021