Extended Abstract

DNA Barcoding und Next Generation Sequencing

Harald Meimberg

Die technologische Entwicklung für die DNA-Sequenzierung schreitet sehr schnell voran, so dass diese Entwicklungen für aktuelle und zukünftige Datensammlungen berücksichtigt werden sollte, insbesondere wenn sie sich langfristig als Referenz eignen soll. Dieser Zusammenhang ergibt sich bei DNA-Barcodedaten, die gesammelt werden, um eine Basis für den Vergleich von Sequenzdaten für unbestimmte Proben, aber auch für Bulk-Proben wie Meta-Barcoding oder *Environmental DNA*. Besonders für die letzteren ist die aktuelle Form der Barcodes nicht ideal.

DNA-Barcodierung als Verwendung eines standardisierten DNA-Sequenz-Referenzsystems zur Artenidentifikation wurde vor etwa 15 Jahren konzeptionell entwickelt. Eine frühzeitige Diskussion über die Anwendbarkeit von DNA-Sequenzen und über das Wie, und Warum, ein standardisiertes Referenzsystem zu etablieren, der Barcodierungsteil der molekularen Systematik, wurde in prominenten Zeitschriften um 2003 (z. B. Tauz et al. 2003) präsentiert. Es folgte direkt aus den Fortschritten der DNA-Sequenzierung nach dem Sanger-Prinzip, das in den 1990er Jahren perfektioniert wurde, mit dem Höhepunkt der Fertigstellung der frühen Genomprojekte zu Beginn dieses Jahrhunderts und aus seiner Anwendungen auf molekulare Phylogenetik. Infolgedessen wurden Methoden und Verfahren vorgeschlagen, die hinsichtlich Output, des Datenformats und der Analyseleistung optimiert sind, die normalerweise durch die Sanger-Sequenzierung erreicht werden können. Für DNA-Barcodes bedeutet dies die Verwendung von Regionen, die zwischen 700 und 1000 bp lang und von konservierten Regionen flankiert sind, um Primer-Design zu ermöglichen. Vor allem in der Zoologie wurde darüber hinaus die Verwendung von einer einzelnen Region als Barcode favorisiert, die mit einer Sanger-Seqeunzierung erfasst werden kann (Herbert et al. 2003).

Der Grund dafür ist, die maximale Menge an Informationen in Bezug auf den Aufwand in Form von Amplifikations- und Sequenzierungsreaktionen zu erreichen. Diese Grundlagen sind bei der Verwendung der heutigen (Next Generation) Sequenztechnologie anders. Die Sequenzlänge beträgt etwa 500 bp, aber neben den Amplicons können Sequenzdaten als *Shot Gun* Sequenzen gesammelt werden, d.h. als zufällige DNA-Fragmente, die bioinformatisch identifiziert werden (Stull et al. 2015, Richardson et al. 2015). Andererseits ist die Erhöhung des Aufwandes für die Sequenzierung zusätzlicher Loci, geringer als bei der Sanger-Sequenzierung, so dass eine höhere Sequenzmenge erzeugt werden kann (Shokralla et al. 2015). Daher geht der Bedarf in Bezug auf Anzahl von Sequenzen, die gesammelt und zusammen mit belegter taxonomischer Informationen gespeichert werden, über die aktuell gesammelten Barcode-Regionen hinaus. Es ist wahrscheinlich, dass dies in der Zukunft noch stärker ausgeprägt ist: Wenn man bedenkt, wie schnell die Datenerfassungskapazität in den letzten 15 Jahren gestiegen ist, können wir auch in den nächsten 15 Jahren eine tiefgehende Veränderung erwarten.

Meimberg H., 2017: DNA Barcoding and Next Generation Sequencing.
Technology for DNA sequencing is developing very fast and it is important to consider these developments for current and future data collections; in particular, if it is

intended to provide a long-term suitability as a reference. In this respect, DNA barcode data are collected to provide a base for comparison for sequence data for undetermined samples but also for bulk samples like meta-barcoding or environmental DNA. Especially for the latter the current form of barcodes might not be ideal.
DNA barcoding as the use of a standardized DNA sequence reference system for species identification had been conceptually developed about 15 years ago. Early discussion about the applicability of DNA sequences and how, and if, a standardized reference system should be established, the barcoding part of molecular systematics, was presented in prominent journals around 2003 (e.g. Tauz et al. 2003). It directly followed the advances of DNA sequencing using the Sanger principle that was perfected during the 1990ies with its peak of the completion of the early genome projects beginning this century, and its applications for molecular phylogenetic analyses. In consequence, methods and procedures were suggested that are optimized in regard to output, data format and analysis power that can be normally attained by Sanger sequencing. For DNA barcodes this means the utilization of regions that are between 700 and 1000 bp long and that are flanked by conserved regions to allow primer design. In addition, especially in zoology, the use of a single region as barcode was preferred, that can be covered by one Sanger sequence (Herbert et al. 2003).
The reason for this is to attempt the maximum amount of information in relation to effort in form of amplification and sequencing reactions. These principles are different when using nowadays available (next generation) sequencing technology. Sequence length are around 500 bp but next to amplicons, sequence data can be collected as shotgun sequences, i.e. as random DNA fragments that are identified bioinformatically (Stull et al. 2015, Richardson et al. 2015). On the other hand, increase of effort if additional loci are sequenced is less than with Sanger sequencing, so a higher amount of sequences can be generated (Shokralla et al. 2015). Thus, the demand of amount of sequences that are collected and saved together with vouchered taxonomic information goes beyond the currently collected barcode regions. It is likely that this is more pronounced in the future: considering how fast data collection capacity increased the last 15 years, we can expect a profound change probably also during the next 15 years.

Keywords: DNA Barcoding, Next Generation Sequencing.

Literatur

Shokralla S., Porter T.M., Gibson J.F., Dobosz R., Janzen D.H., Hallwachs W., Golding G.B. & Hajibabaei M., 2015: Massively parallel multiplex DNA sequencing for specimen identification using an Illumina MiSeq platform. Sci. Rep. 5. doi:10.1038/srep09687.

Stull G.W., Moore M.J., Mandala V.S., Douglas N.A., Kates H-R., Qi X., Brockington S.F., Soltis P.S., Soltis D.E. & Gitzendanner M.A., 2013: A targeted enrichment strategy for massively parallel sequencing of angiosperm plastid genomes. Appl. Plant Sci. 1(2). doi:10.3732/apps.1200497.

Richardson R.T., Lin C-H., Quijia J.O., Riusech N.S., Goodell K., & Johnson R.M., 2015: Rank-based characterization of pollen assemblages collected by honey bees using a multi-locus metabarcoding approach. Appl. Plant Sci. 3(11). doi: 10.3732/apps.1500043.

Tautz D., Arctander P., Minelli A., Thomas R.H., & Vogler A.P., 2003: A plea for DNA taxonomy. Trends Ecol. Evol. 18(2), 70–74.

Hebert P.D.N., Ratnasingham S. & de Waard J.R., 2003: Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species (figure 1). Proc. R. Soc. Lond., B, Biol. Sci. 270, 96–99. doi:10.1098/rsbl.2003.0025.

Anschrift:

Univ.-Prof. Dr. Harald Meimberg, Institute for Integrative Nature Conservation Research, University for Life Sciences and Natural Resources (Boku), Vienna.
E-Mail: meimberg@boku.ac.at

ZOBODAT - www.zobodat.at

Zoologisch-Botanische Datenbank/Zoological-Botanical Database

Digitale Literatur/Digital Literature

Zeitschrift/Journal: Verhandlungen der Zoologisch-Botanischen Gesellschaft in Wien. Frueher: Verh.des Zoologisch-Botanischen Vereins in Wien. seit 2014 "Acta ZooBot Austria"

Jahr/Year: 2017

Band/Volume: 154

Autor(en)/Author(s): Meimberg Harald

Artikel/Article: DNA Barcoding und Next Generation Sequencing 185-186