



Vergabeverfahren UMR-2026-0011

Leistungsbeschreibung

DNA-Barcoding, DNA-Metabarcoding und Mitogenomsequenzierung von Käfern (Coleoptera)

1. Gegenstand und Zielsetzung der Ausschreibung

Gegenstand der Ausschreibung ist die umfassende molekulargenetische Untersuchung von Insektenproben, insbesondere Käfern (Coleoptera), im Rahmen eines wissenschaftlichen Drittmittelprojekts. Die Untersuchungen dienen der taxonomischen Identifikation, der Erfassung der Artenzusammensetzung sowie der Analyse genetischer und phylogenetischer Verwandtschaftsverhältnisse.

Ein besonderer Schwerpunkt liegt auf der standardisierten, reproduzierbaren und langfristig vergleichbaren Aufbereitung molekulargenetischer Daten. Insbesondere für Los I sollen die Ergebnisse eine projektübergreifende Integration und Harmonisierung neuer und historischer Sequenzdatensätze innerhalb einer cloudbasierten Dateninfrastruktur ermöglichen.

Ziel ist die Schaffung einer wissenschaftlich belastbaren und nachhaltig nutzbaren Datengrundlage.

Die ausgeschriebenen Leistungen gliedern sich in zwei Lose:

Los I – DNA-Metabarcoding

Molekulare Erfassung der Artenzusammensetzung von 400 Insekten-Mischproben mittels COI-basierter Hochdurchsatzsequenzierung einschließlich bioinformatischer Artidentifikation, taxonomischer Harmonisierung sowie ökologischer und naturschutzfachlicher Charakterisierung der nachgewiesenen Taxa.

Los II – DNA-Barcoding und Mitogenomsequenzierung

Umfassende molekulargenetische Charakterisierung von bis zu 120 Einzelindividuen durch die Sequenzierung und Rekonstruktion vollständiger mitochondrialer Genome sowie die ergänzende Analyse von bis zu zehn nukleären Genregionen. Die Auswahl dieser nukleären Marker erfolgt in Abstimmung mit dem Auftraggeber. Die gewonnenen Sequenzdaten dienen insbesondere der vergleichenden Genomanalyse und der Rekonstruktion molekularer Phylogenien.



Die Leistungen umfassen sämtliche erforderlichen Schritte der Probenaufbereitung, DNA-Extraktion, PCR-Amplifikation, Sequenzierung und bioinformatischen Auswertung sowie die qualitätsgesicherte Bereitstellung und Dokumentation der Ergebnisse.

Los I – DNA-Metabarcoding von Insekten-Mischproben

2. Probenaufbereitung und DNA-Extraktion

Der Auftragnehmer führt für insgesamt 400 Insekten-Mischproben eine standardisierte DNA-Metabarcoding-Analyse durch, die eine möglichst umfassende Erfassung und taxonomische Identifikation der enthaltenen Insektentaxa ermöglicht.

Sämtliche Insekten-Mischproben sind jeweils als Gesamtprobe ohne vorherige Größenfraktionierung aufzubereiten. Hierzu ist das Probenmaterial nach einem standardisierten und qualitätsgesicherten Verfahren zu homogenisieren und anschließend einer DNA-Extraktion zu unterziehen.

Die Probenaufbereitung und DNA-Extraktion müssen so erfolgen, dass eine möglichst repräsentative Erfassung der enthaltenen Insektentaxa gewährleistet wird. Geeignete Maßnahmen zur Qualitätssicherung und zur Vermeidung von Kreuzkontaminationen sind vorzusehen.

3. PCR-Amplifikation und Hochdurchsatzsequenzierung

Für die DNA-Metabarcoding-Analyse ist der mitochondriale Marker Cytochrom-c-Oxidase I (COI) unter Verwendung des etablierten Primeransatzes nach Leray et al. (2013) einzusetzen. Der Primeransatz muss die Amplifikation eines etwa 313 bp langen COI-Fragments ermöglichen.

Die PCR-Amplifikation, Library-Präparation und anschließende Hochdurchsatzsequenzierung erfolgen nach einem standardisierten und qualitätsgesicherten Verfahren. Geeignete Negativkontrollen sind in die Laborprozesse einzubeziehen.

Die Hochdurchsatzsequenzierung ist mit einer Zieltiefe von mindestens 250.000 qualitätsgeprüften Paired-End-Readpaaren je Insekten-Mischprobe durchzuführen. Für die insgesamt 400 Insekten-Mischproben ergibt sich damit eine angestrebte Gesamttiefe von mindestens 75 Millionen qualitätsgeprüften Paired-End-Readpaaren.

Das Angebot muss Angaben zur vorgesehenen Sequenzierungsplattform, zur Library-Präparation, zur Indexierungsstrategie und zur geplanten Sequenzierkapazität enthalten.

4. Bioinformatische Auswertung und Qualitätssicherung

Die gewonnenen Sequenzdaten sind mittels einer standardisierten, qualitätsgesicherten und reproduzierbaren bioinformatischen Pipeline auszuwerten.



Die Auswertung umfasst mindestens:

- Qualitätskontrolle und Filterung der Rohsequenzen einschließlich der Entfernung von Primern und technischen Sequenzbestandteilen
- Zusammenführung geeigneter Paired-End-Sequenzen und Entfernung minderwertiger beziehungsweise artefaktbehafteter Sequenzen
- Identifikation und Gruppierung geeigneter Sequenzvarianten, beispielsweise anhand von Amplicon Sequence Variants (ASVs), Exact Sequence Variants (ESVs) oder vergleichbaren Sequenzclustern (OTUs)
- Taxonomische Identifikation der nachgewiesenen Sequenzvarianten anhand dokumentierter Referenzdatenbanken
- Erstellung einer nach Proben aufgeschlüsselten Übersicht der nachgewiesenen Sequenzvarianten und ihrer absoluten sowie relativen Readhäufigkeiten

Die eingesetzten Qualitätsfilter, Sequenzschwellenwerte, Softwareversionen und Referenzdatenbanken sind nachvollziehbar zu dokumentieren.

5. Taxonomische Referenzierung anhand mehrerer Datenbanken

Für die taxonomische Identifikation ist ein paralleler Abgleich gegen mindestens drei Referenzdatenbanken durchzuführen:

- BOLD (Barcode of Life Data System)
- NCBI GenBank (National Center for Biotechnology Information)
- MIDORI2 (kuratierte mitochondriale Referenzsequenzdatenbank)

Die Ergebnisse der einzelnen Datenbankabgleiche sind vergleichend auszuwerten. Übereinstimmungen, widersprüchliche taxonomische Zuordnungen sowie Unterschiede in der taxonomischen Auflösung müssen nachvollziehbar dokumentiert werden.

Die Zuverlässigkeit der taxonomischen Identifikation ist unter Berücksichtigung geeigneter Sequenzähnlichkeitswerte, der Qualität der Referenzsequenzen und der Übereinstimmung zwischen den verwendeten Referenzdatenbanken zu bewerten.

Für die jeweiligen Sequenzvarianten müssen die wesentlichen Referenztreffer einschließlich Referenzkennung, taxonomischer Identifikation und Sequenzähnlichkeitswerten exportierbar sein.

6. Taxonomische Harmonisierung und naturschutzfachliche Bewertung

Die identifizierten Taxa sind zusätzlich mit der GBIF Backbone Taxonomy abzugleichen, um eine einheitliche taxonomische Referenzierung hinweg zu gewährleisten.

Hierbei sind insbesondere akzeptierte wissenschaftliche Artnamen, Synonyme und verfügbare eindeutige taxonomische Identifikatoren zu berücksichtigen.

Darüber hinaus ist ein Abgleich mit den jeweils aktuellen beziehungsweise zuletzt veröffentlichten Roten Listen Deutschlands für die relevanten Insektengruppen durchzuführen.

Für die nachgewiesenen Taxa sind, soweit verfügbar und taxonomisch hinreichend sicher zuordenbar, die jeweiligen Gefährdungskategorien, Referenzquellen und Bewertungsstände auszuweisen.

Die Ergebnisse müssen eine Identifikation naturschutzfachlich relevanter, insbesondere gefährdeter Insektenarten ermöglichen.

7. Berücksichtigung von Barcode Index Numbers (BINs) und BIN-Sharing

Im Rahmen der taxonomischen Identifikation sind verfügbare Barcode Index Numbers (BINs) aus BOLD zu berücksichtigen und den nachgewiesenen Sequenzvarianten zuzuordnen, soweit dies anhand geeigneter Referenzsequenzen möglich ist.

Insbesondere ist zu prüfen, ob einzelne BINs in BOLD mit mehreren unterschiedlichen Artnamen verknüpft sind (BIN-Sharing).

Für jede entsprechend identifizierte BIN ist eine Übersicht sämtlicher öffentlich verfügbarer, auf Artniveau bestimmter Referenzdatensätze innerhalb dieser BIN bereitzustellen. Diese muss mindestens folgende Informationen enthalten:

- Eindeutige BIN-Kennung (BOLD BIN URI)
- Anzahl unterschiedlicher innerhalb der BIN geführter Arten
- Vollständige Auflistung aller zugeordneten Artnamen
- Anzahl der in BOLD hinterlegten, auf die jeweilige Art bestimmten Referenzdatensätze innerhalb der BIN
- Kennzeichnung taxonomisch mehrdeutiger oder nicht eindeutig auf Artniveau auflösbarer Zuordnungen

Sofern mehrere Arten innerhalb einer BIN geführt werden, darf eine eindeutige Artidentifikation nicht allein aufgrund der BIN-Zugehörigkeit erfolgen. Stattdessen sind sämtliche möglichen Arten einschließlich ihrer jeweiligen Referenzhäufigkeiten transparent darzustellen.

Bei Verwendung des COI-Mini-Barcodes nach Leray et al. (2013) kann die BIN-Zuordnung indirekt über geeignete BOLD-Referenzsequenzen erfolgen. Die Unsicherheit einer solchen Zuordnung ist entsprechend zu kennzeichnen.

8. Cloubasierte Datenharmonisierung und Integration historischer Datensätze

Die bioinformatische Verarbeitung muss so erfolgen, dass die resultierenden Sequenzvarianten, Sequenzcluster und taxonomischen Zuordnungen innerhalb einer cloubasierten Dateninfrastruktur projektübergreifend gespeichert, verglichen und harmonisiert werden können.

Insbesondere muss die Möglichkeit bestehen, neu generierte Metabarcoding-Daten mit bereits vorhandenen historischen Sequenzdatensätzen (Legacy-Daten) zusammenzuführen und über einheitliche Sequenzcluster beziehungsweise harmonisierte taxonomische Referenzierungen vergleichbar zu machen.

Die Datenharmonisierung muss auf Sequenzebene erfolgen, um eine langfristige Vergleichbarkeit unabhängig von Änderungen der taxonomischen Nomenklatur und der jeweils verwendeten Referenzdatenbanken zu ermöglichen.

Die cloudbasierte Auswertungsumgebung muss darüber hinaus eine regelmäßige Aktualisierung taxonomischer Zuordnungen anhand fortlaufend erweiterter Referenzdatenbanken ermöglichen.

Änderungen der taxonomischen Zuordnungen sollen sowohl für bestehende als auch für neu hinzukommende Datensätze nachvollziehbar berücksichtigt werden können, ohne dass hierfür eine erneute molekularbiologische Analyse erforderlich ist.

9. Ergebnisbereitstellung für Los I

Die Ergebnisse sind in einer standardisierten, maschinenlesbaren und nachvollziehbar dokumentierten Form bereitzustellen.

Die Lieferung umfasst verpflichtend:

- Vollständige Sequenzrohdaten im FASTQ-Format, einschließlich der demultiplexierten Read-Dateien je Probe
- Qualitätsberichte zur Sequenzierung und bioinformatischen Verarbeitung
- FASTA-Dateien der resultierenden Sequenzvarianten beziehungsweise Sequenzcluster
- Nachweismatrix der Sequenzvarianten und ihrer Readhäufigkeiten je Probe
- Taxonomische Identifikationen und Referenztreffer aus BOLD, NCBI GenBank und MIDORI2
- GBIF-taxonomische Zuordnungen und verfügbare Rote-Liste-Kategorien
- BIN-Zuordnungen einschließlich vollständiger Auflistung möglicher Arten und ihrer jeweiligen Referenzhäufigkeiten
- Dokumentation sämtlicher wesentlicher Labor- und Bioinformatikverfahren sowie Qualitätsparameter

Die Ergebnisse müssen zusätzlich über eine cloudbasierte Analyse- und Visualisierungsumgebung zugänglich sein, die eine interaktive Exploration der taxonomischen und ökologischen Ergebnisse sowie eine projektübergreifende vergleichende Analyse ermöglicht.

Los II – DNA-Barcoding, Mitogenomsequenzierung und phylogenetische Analyse

10. Probenaufbereitung und DNA-Extraktion

Der Auftragnehmer führt für insgesamt bis zu 120 individuell vorliegende Organismenproben eine molekulargenetische Analyse durch, die sowohl die vollständige Sequenzierung der mitochondrialen Genome als auch die gezielte Untersuchung ausgewählter nukleärer Genregionen umfasst.

Hierzu ist aus sämtlichen Individuen genomische DNA mittels eines geeigneten, standardisierten und qualitätsgesicherten Verfahrens zu extrahieren.

Die Probenaufbereitung und DNA-Extraktion müssen unter kontrollierten Laborbedingungen mit räumlich oder organisatorisch getrennten Arbeitsbereichen für kontaminationskritische Arbeitsschritte erfolgen.

Die DNA-Qualität und -Konzentration sind hinsichtlich ihrer Eignung für die nachfolgenden Analysen zu überprüfen.

11. Etablierung und Optimierung der Analysestrategie

Vor Beginn der vollständigen Probenbearbeitung sind in enger Abstimmung mit dem Auftraggeber initiale Machbarkeits- und Optimierungstests an einer repräsentativen Auswahl weniger Individuen durchzuführen.

Diese dienen insbesondere der Auswahl und Validierung einer geeigneten Sequenzierungs- und Auswertungsstrategie für die Rekonstruktion vollständiger Mitogenome sowie der Überprüfung der Amplifizierbarkeit der vorgesehenen nukleären Genregionen.

Das Angebot muss eine Beschreibung des vorgesehenen methodischen Ansatzes einschließlich Sequenzierungstechnologie, Library-Präparation und bioinformatischer Auswertungsstrategie enthalten.

12. Vollständige Sequenzierung und Rekonstruktion mitochondrialer Genome

Für jedes Individuum ist die möglichst vollständige Rekonstruktion des mitochondrialen Genoms anzustreben.

Hierzu ist ein geeignetes Hochdurchsatzsequenzierungsverfahren einzusetzen, das die Erfassung der mitochondrialen Genregionen einschließlich protein-kodierender Gene, ribosomaler RNA-Gene und Transfer-RNA-Gene ermöglicht.

Die Analyse umfasst insbesondere COI, COII, COIII, Cytochrom b, 12S, 16S, NADH-Dehydrogenase-Untereinheiten sowie weitere mitochondriale Genbereiche.

Die gewonnenen Sequenzdaten sind mittels geeigneter bioinformatischer Verfahren zu assemblieren und, soweit möglich, zu vollständigen mitochondrialen Genomsequenzen zu rekonstruieren.

Hierbei sind sowohl de-novo-Assemblierungsverfahren als auch referenzgestützte Ansätze zulässig.

Für jedes Individuum sind die Qualität und Vollständigkeit des rekonstruierten Mitogenoms zu bewerten und zu dokumentieren. Dies umfasst insbesondere Sequenzabdeckung, Sequenziertiefe, Vollständigkeit der erwarteten mitochondrialen Genregionen und mögliche Assemblierungsunsicherheiten.

13. Referenz-Mapping und vergleichende Mitogenomanalyse

Die rekonstruierten mitochondrialen Genomsequenzen sind gegen ein geeignetes Referenzpanel phylogenetisch nah verwandter Arten und Gattungen abzugleichen. Das Referenzpanel ist in Abstimmung mit dem Auftraggeber zusammenzustellen und muss geeignete, öffentlich verfügbare mitochondriale Referenzgenome aus anerkannten Sequenzdatenbanken, insbesondere NCBI GenBank beziehungsweise RefSeq, berücksichtigen.

Die vergleichende Analyse umfasst mindestens:

- Auswahl und Dokumentation geeigneter Referenzgenome nah verwandter Arten und Gattungen
- Mapping der gewonnenen Sequenzdaten gegen ausgewählte mitochondriale Referenzgenome
- Bewertung von Sequenzähnlichkeit und Genomabdeckung
- Identifikation homologer mitochondrialer Genregionen und Erstellung geeigneter Sequenzalignments
- Bereitstellung konsolidierter mitochondrialer Genomsequenzen und homologer Genalignments für weiterführende phylogenetische Analysen

Die resultierenden Sequenzdaten sind in standardisierten, phylogenetisch weiterverarbeitbaren Formaten, insbesondere FASTA sowie geeigneten Alignmentformaten wie PHYLIP oder NEXUS, zu exportieren.

Das verwendete Referenzpanel muss einschließlich sämtlicher Referenzsequenzen, Accession-Nummern und taxonomischer Metadaten mit den Ergebnissen bereitgestellt werden, soweit dies die jeweiligen Datenbank- und Nutzungsbedingungen zulassen.

14. Ergänzendes DNA-Barcoding nukleärer Genregionen

Ergänzend zur mitochondrialen Genomsequenzierung sind für jedes Individuum bis zu zehn ausgewählte nukleäre Genregionen (nDNA) mittels gezielter PCR-Amplifikation und anschließender DNA-Sequenzierung zu untersuchen.

Die Auswahl und Festlegung der bis zu zehn nukleären Marker erfolgt in enger Abstimmung mit dem Auftraggeber und orientiert sich an deren Eignung zur phylogenetischen Differenzierung der untersuchten Taxa.

Hierbei können insbesondere etablierte nukleäre Marker wie 18S rRNA, 28S rRNA sowie weitere geeignete protein-kodierende oder nicht-kodierende Genregionen berücksichtigt werden.

Die Analyse umfasst die Etablierung beziehungsweise Anpassung geeigneter PCR-Assays, die Amplifikation der abgestimmten Zielregionen, die Sequenzierung der PCR-Produkte sowie deren qualitätsgesicherte bioinformatische Auswertung.

Die resultierenden nukleären Sequenzen sind je Individuum und Genregion eindeutig zuzuordnen und in standardisierten Sequenzformaten bereitzustellen.

Bei Durchführung von Sanger-Sequenzierungen müssen die vollständigen originalen Elektropherogrammdateien im ABI-Format (.ab1) verpflichtend mitgeliefert werden. Eine ausschließliche Bereitstellung der daraus abgeleiteten FASTA-Konsensussequenzen ist nicht ausreichend.

Bei Verwendung eines Hochdurchsatzsequenzierungsverfahrens sind stattdessen die zugehörigen Rohsequenzdaten im FASTQ-Format sowie die resultierenden Konsensussequenzen bereitzustellen.

Für sämtliche nukleären Genregionen sind geeignete Sequenzalignments zu erstellen, die für eine eigenständige oder kombinierte phylogenetische Auswertung mit mitochondrialen Sequenzdaten genutzt werden können.

15. Bioinformatische Qualitätssicherung und Datenaufbereitung

Die Sequenzrohdaten sind einer standardisierten bioinformatischen Qualitätskontrolle und Aufbereitung zu unterziehen.

Diese umfasst mindestens das Demultiplexing, die Entfernung technischer Adaptersequenzen, die Qualitätsfilterung und die Bewertung der gewonnenen Sequenzdaten.

Für die mitochondrialen Genome sind geeignete Verfahren zur Assemblierung, Referenzzuordnung und Annotation anzuwenden.

Mögliche Kontaminationen, Assemblierungsartefakte und nukleäre mitochondriale Sequenzkopien (NUMTs) sind soweit methodisch möglich zu erkennen und bei der Interpretation zu berücksichtigen.

Sämtliche Sequenzdaten müssen den jeweiligen Individuen eindeutig zuordenbar sein und eine reproduzierbare Weiterverarbeitung ermöglichen.

16. Phylogenetische Auswertung

Auf Grundlage der rekonstruierten mitochondrialen Genome und der ergänzend sequenzierten nukleären Genregionen ist eine molekulare Phylogenie der untersuchten Individuen beziehungsweise Arten zu erstellen.

Hierzu sind geeignete homologe Genregionen auszuwählen, Sequenzalignments zu erzeugen und mit den abgestimmten Referenztaxa zu kombinieren.

Die phylogenetische Rekonstruktion muss nach einem etablierten Verfahren, beispielsweise mittels Maximum-Likelihood-Analyse, erfolgen.

Die statistische Unterstützung der resultierenden Verzweigungen ist durch geeignete Verfahren, etwa Bootstrap-Analysen, zu bewerten.

Soweit fachlich sinnvoll, sollen mitochondriale und nukleäre Datensätze sowohl getrennt als auch in einer kombinierten Analyse berücksichtigt werden.

Die resultierenden phylogenetischen Stammbäume sind einschließlich ihrer statistischen Unterstützungswerte in geeigneten Grafikformaten und maschinenlesbaren Formaten, insbesondere Newick, bereitzustellen.

17. Wiederholungsanalysen und Qualitätssicherung

Das Angebot muss Angaben zu den vorgesehenen Qualitätssicherungsmaßnahmen sowie zum Umgang mit fehlgeschlagenen DNA-Extraktionen, PCR-Amplifikationen oder Sequenzierungsreaktionen enthalten.

Insbesondere ist darzustellen, unter welchen Voraussetzungen kostenfreie Wiederholungen fehlgeschlagener Amplifikationsschritte erfolgen, sofern bei ausreichender DNA-Qualität und -Menge im ersten Durchgang keine für die weitere Analyse geeigneten Ergebnisse erzielt werden konnten.

Abweichungen vom vorgesehenen Analyseumfang, insbesondere unvollständige Mitogenomrekonstruktionen oder nicht erfolgreich sequenzierte nukleäre Marker, sind nachvollziehbar zu dokumentieren und mit dem Auftraggeber abzustimmen.

18. Ergebnisbereitstellung für Los II

Die Ergebnisbereitstellung umfasst verpflichtend:

- Vollständige Sequenzrohdaten im FASTQ-Format bei Hochdurchsatzsequenzierung
- Originale ABI-Dateien (.ab1) sämtlicher Sanger-Sequenzierungen, einschließlich verfügbarer Vorwärts- und Rückwärtssequenzierungen
- Assemblierte und annotierte mitochondriale Genomsequenzen je Individuum im FASTA-Format, soweit erfolgreich rekonstruierbar
- Dokumentation der Vollständigkeit, Sequenzabdeckung und Qualität der Mitogenomassemblierungen
- Verwendete mitochondriale Referenzgenome einschließlich Accession-Nummern und taxonomischer Metadaten
- Ergebnisse des Referenz-Mappings
- Sequenzdaten der bis zu zehn untersuchten nukleären Genregionen je Individuum
- Multiple Sequenzalignments der mitochondrialen und nukleären Genregionen
- Phylogenetische Stammbäume einschließlich statistischer Unterstützungswerte in grafischer Darstellung und maschinenlesbarem Format
- Qualitätsberichte und einen zusammenfassenden Ergebnisbericht einschließlich Methodik, Qualitätssicherung und phylogenetischer Interpretation

Die Bereitstellung ausschließlich bereits bearbeiteter Konsensussequenzen ohne die zugehörigen verfügbaren Rohsequenzdaten ist nicht zulässig.

Verpflichtend mit dem Angebot einzureichende Unterlagen

19. Labor- und Analyseplan

Mit dem Angebot ist für jedes angebotene Los ein detaillierter Labor- und Analyseplan einzureichen.

Der Labor- und Analyseplan muss die vorgesehene praktische Durchführung der Leistung nachvollziehbar beschreiben und mindestens folgende Angaben enthalten:

- Probenannahme, Probenregistrierung und Probenaufbereitung
- Verfahren zur DNA-Extraktion und eingesetzte Extraktionsmethoden
- Vorgesehene PCR-Marker, Primer und grundsätzliche Amplifikationsstrategie
- Sequenzierungsplattform, Library-Präparation und Sequenzierungstechnologie
- Geplante Sequenziertiefe beziehungsweise Sequenzabdeckung
- Qualitätssicherungsmaßnahmen einschließlich Negativkontrollen
- Bioinformatische Verarbeitung einschließlich verwendeter Software, Qualitätsfilter und Referenzdatenbanken
- Vorgesehene Ergebnisformate und Datentransferverfahren
- Zeitlicher Ablauf der wesentlichen Arbeitsschritte einschließlich der vorgesehenen Abstimmungen mit dem Auftraggeber

Für Los II sind darüber hinaus die geplante Mitogenom-Assemblierungsstrategie, der Aufbau des Referenzpanels, die Vorgehensweise zur Analyse der nukleären Marker sowie die Methoden der phylogenetischen Rekonstruktion darzustellen.

20. Verbindliche Anforderungen an die Rohdatenlieferung

Für beide Lose gilt, dass sämtliche im Rahmen des Auftrags generierten und für die wissenschaftliche Nachnutzung erforderlichen Sequenzrohdaten dem Auftraggeber vollständig zur Verfügung zu stellen sind.

Hierbei gelten insbesondere folgende Anforderungen:

- FASTQ: Vollständige Rohsequenzdaten sämtlicher Hochdurchsatzsequenzierungen, einschließlich der jeweiligen Qualitätswerte und eindeutiger Probenzuordnung
- ABI (.ab1): Vollständige originale Elektropherogrammdateien sämtlicher im Auftrag durchgeführter Sanger-Sequenzierungen
- FASTA: Sämtliche erzeugten Konsensussequenzen, Sequenzvarianten, Sequenzcluster und assemblierten Genomsequenzen



- Metadaten: Eindeutige Zuordnung sämtlicher Dateien zu Probe, Individuum, Marker und gegebenenfalls Referenzsequenz
- Referenzinformationen: Dokumentation der eingesetzten Referenzdatenbanken, Referenzsequenzen, Accession-Nummern beziehungsweise taxonomischen Identifikatoren sowie der verwendeten Datenbankversionen oder Abrufstände

Die Rohsequenzdaten sind in ihren ursprünglichen Qualitätsinformationen erhaltenden Formaten bereitzustellen. Die Bereitstellung ausschließlich bereits gefilterter, zusammengeführter oder taxonomisch zugeordneter Sequenzen ist nicht ausreichend.

Die Lieferung der Rohdaten und sämtlicher vereinbarter Ergebnisdateien ist Bestandteil der ausgeschriebenen Leistung und darf nicht von einer zusätzlichen Vergütung abhängig gemacht werden.

21. Vollständigkeit der Angebotsunterlagen

Die Vorlage des Labor- und Analyseplans, der fachlichen Referenzen sowie der repräsentativen Beispieldaten ist Bestandteil der geforderten Angebotsunterlagen.

Die Unterlagen müssen eine nachvollziehbare Bewertung der vorgesehenen Laborverfahren, der vorhandenen bioinformatischen Leistungsfähigkeit, der wissenschaftlichen Erfahrung sowie der Qualität und Nutzbarkeit der angebotenen Ergebnisse ermöglichen.

Das Angebot muss darüber hinaus erkennen lassen, dass die geforderten Sequenzrohdaten, insbesondere FASTQ-Dateien und bei Sanger-Sequenzierung ABI-Dateien (.ab1), im Rahmen der Leistungserbringung vollständig bereitgestellt werden können.