

Analyse von RAD-Seq-Daten unter Berücksichtigung von Sequenzierfehlerraten und Heterozygotiewahrscheinlichkeiten

Antonie Vietor

24. Februar 2021

Technische Universität Dortmund
Fakultät für Informatik
Lehrstuhl 11
Bioinformatics for High-Throughput Technologies
<http://ls11-www.cs.tu-dortmund.de/>

In Kooperation mit:
Universität Duisburg-Essen
Genome Informatics
<http://genomeinformatics.uni-due.de/>

Aufbau von DNA und RNA

Aufbau der DNA

- besteht aus Nukleotiden
- jedes **Nukleotid** besteht aus einem Zuckermolekül (Desoxyribose), einem Phosphatrest und einer Base
- **Basen**: A (Adenin), T (Thymin), G (Guanin), C (Cytosin)
- meist **doppelsträngig**
- dient vor allem der **Informationsspeicherung** (Erbinformation)

Aufbau von DNA und RNA

Aufbau der DNA

- besteht aus Nukleotiden
- jedes **Nukleotid** besteht aus einem Zuckermolekül (Desoxyribose), einem Phosphatrest und einer Base
- **Basen**: A (Adenin), T (Thymin), G (Guanin), C (Cytosin)
- meist **doppelsträngig**
- dient vor allem der **Informationsspeicherung** (Erbinformation)

Unterschiede im Aufbau der RNA

- **Nukleotide**: das Zuckermolekül ist Ribose
- **Basen**: Uracil (U) statt Thymin
- meist **einzelsträngig**
- viele Funktionen, dient unter anderem der **Informationsübertragung** bei der Proteinbiosynthese

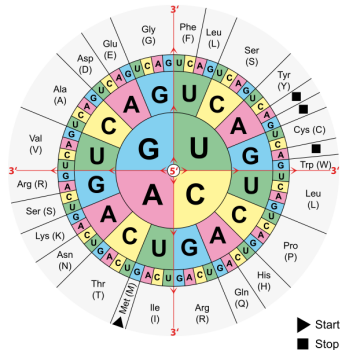
Struktur der DNA

- **Doppelhelixstruktur**
- **Komplementarität:** selektive Basenpaarung von A und T und ebenso von G und C
- **Antiparallelität:** in der Doppelhelix sind die beiden DNA-Stränge gegenläufig zu einander
- **Gene:** Wechsel von codierenden (Exons) und nicht-codierenden Abschnitten (Introns)
- zwischen den Genen nicht-codierende Bereiche, z.T. mit regulatorischen Funktionen
- ca. 98 % der DNA sind nicht-codierend

Proteinbiosynthese

Genetischer Code

- Codierung der **DNA-Sequenz** in eine **Aminosäuresequenz**, welche die Primärstruktur der Proteine darstellt
- **Basentriplets** (Codons) codieren für i.d.R. 20 Aminosäuren sowie ein Start- und drei Stop-Codons
- **Degeneration**: mehrere Basentriplets können für die gleiche Aminosäure codieren

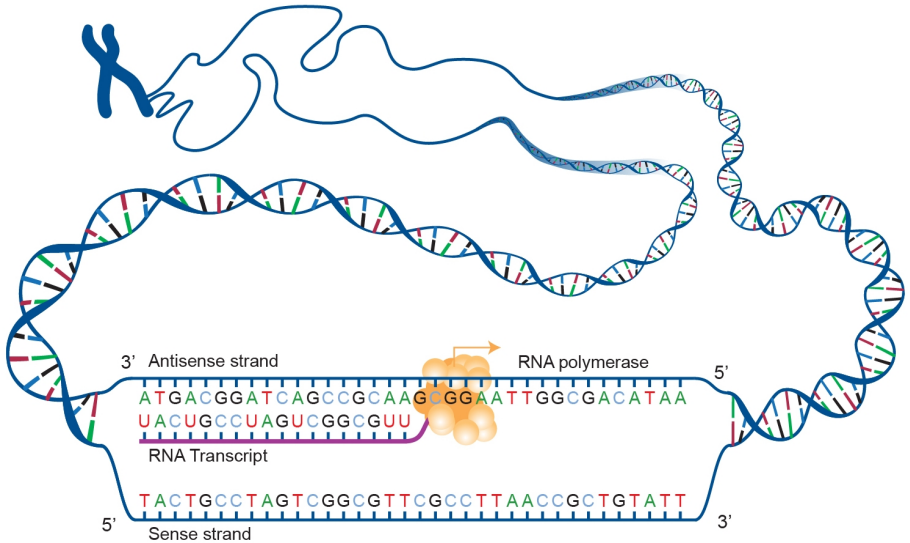


Bildquelle: [1]

Transkription

Umschreiben eines DNA-Abschnitts zu Arbeitskopien in Form von **mRNA** (messenger RNA)

Proteinbiosynthese



Bildquelle: [2]

Translation

- **Übersetzen** der Basensequenz in die Aminosäuresequenz mit Hilfe von **tRNA** (transfer RNA)

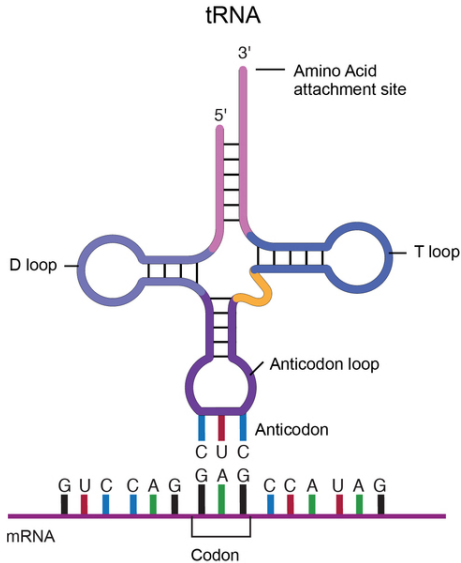
Translation

- **Übersetzen** der Basensequenz in die Aminosäuresequenz mit Hilfe von **tRNA** (transfer RNA)
- Aufbau der tRNA:
 - ⇒ **mRNA-Bindungsstelle** bestehend aus einem Basentriplett

Translation

- **Übersetzen** der Basensequenz in die Aminosäuresequenz mit Hilfe von **tRNA** (transfer RNA)
- Aufbau der tRNA:
 - ⇒ **mRNA-Bindungsstelle** bestehend aus einem Basentriplett
 - ⇒ trägt die **korrespondierende Aminosäure (AS)**, die nach dem genetischen Code der mRNA-Bindungsstelle entspricht

Proteinbiosynthese



Bildquelle: [3]

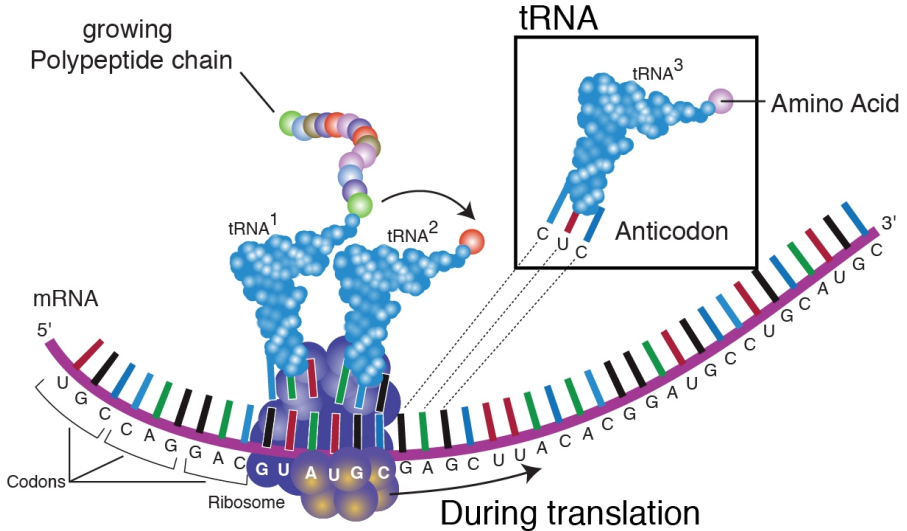
Translation

- **Übersetzen** der Basensequenz in die Aminosäuresequenz mit Hilfe von **tRNA** (transfer RNA)
- Aufbau der tRNA:
 - ⇒ **mRNA-Bindungsstelle** bestehend aus einem Basentriplett
 - ⇒ trägt die **korrespondierende Aminosäure (AS)**, die nach dem genetischen Code der mRNA-Bindungsstelle entspricht

Translation

- **Übersetzen** der Basensequenz in die Aminosäuresequenz mit Hilfe von **tRNA** (transfer RNA)
- Aufbau der tRNA:
 - ⇒ **mRNA-Bindungsstelle** bestehend aus einem Basentriplett
 - ⇒ trägt die **korrespondierende Aminosäure** (AS), die nach dem genetischen Code der mRNA-Bindungsstelle entspricht
- von der Startsequenz ausgehend werden die tRNAs mit komplementärer Bindungsstelle nacheinander an die mRNA gebunden, dadurch wird ihre AS gelöst und an die AS der nachfolgenden tRNA gebunden ⇒ es entsteht eine **Aminosäuresequenz**

Proteinbiosynthese



Bildquelle: [4]

DNA-Replikation

Natürlicher Vorgang zur Vervielfältigung der DNA bei der Zellteilung:

- 1 Entwindung der DNA (**Topoisomerasen**)

DNA-Replikation

Natürlicher Vorgang zur Vervielfältigung der DNA bei der Zellteilung:

- ① Entwindung der DNA (**Topoisomerasen**)
- ② Auftrennung des DNA-Doppelstrangs (**Helikase**)

DNA-Replikation

Natürlicher Vorgang zur Vervielfältigung der DNA bei der Zellteilung:

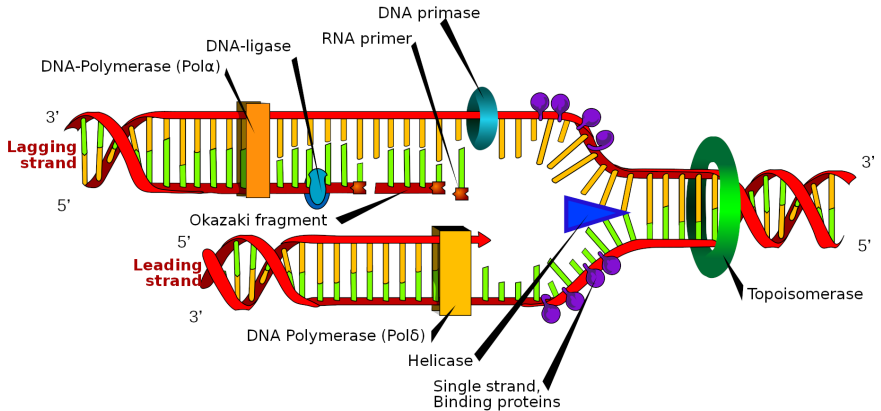
- ① Entwindung der DNA (**Topoisomerasen**)
- ② Auftrennung des DNA-Doppelstrangs (**Helikase**)
- ③ Synthese der RNA-Primer (**Primasen**)

DNA-Replikation

Natürlicher Vorgang zur Vervielfältigung der DNA bei der Zellteilung:

- ① Entwindung der DNA (**Topoisomerasen**)
- ② Auftrennung des DNA-Doppelstrangs (**Helikase**)
- ③ Synthese der RNA-Primer (**Primasen**)
- ④ Kopieren der beiden Elternstränge ausgehend von den RNA-Primern (**DNA-Polymerasen**)

DNA-Replikation



Bildquelle: [5]

DNA-Replikation

Natürlicher Vorgang zur Vervielfältigung der DNA bei der Zellteilung:

- ① Entwindung der DNA (**Topoisomerasen**)
- ② Auftrennung des DNA-Doppelstrangs (**Helikase**)
- ③ Synthese der RNA-Primer (**Primasen**)
- ④ Kopieren der beiden Elternstränge ausgehend von den RNA-Primern (**DNA-Polymerasen**)

DNA-Replikation

Natürlicher Vorgang zur Vervielfältigung der DNA bei der Zellteilung:

- 1 Entwindung der DNA (**Topoisomerasen**)
- 2 Auftrennung des DNA-Doppelstrangs (**Helikase**)
- 3 Synthese der RNA-Primer (**Primasen**)
- 4 Kopieren der beiden Elternstränge ausgehend von den RNA-Primern (**DNA-Polymerasen**)

⇒ es entstehen zwei komplementäre Tochterstränge

DNA-Replikation

Natürlicher Vorgang zur Vervielfältigung der DNA bei der Zellteilung:

- ① Entwindung der DNA (**Topoisomerasen**)
- ② Auftrennung des DNA-Doppelstrangs (**Helikase**)
- ③ Synthese der RNA-Primer (**Primasen**)
- ④ Kopieren der beiden Elternstränge ausgehend von den RNA-Primern (**DNA-Polymerasen**)
 - ⇒ es entstehen zwei komplementäre Tochterstränge
 - ⇒ kontinuierliche Synthese des Leitstrangs

DNA-Replikation

Natürlicher Vorgang zur Vervielfältigung der DNA bei der Zellteilung:

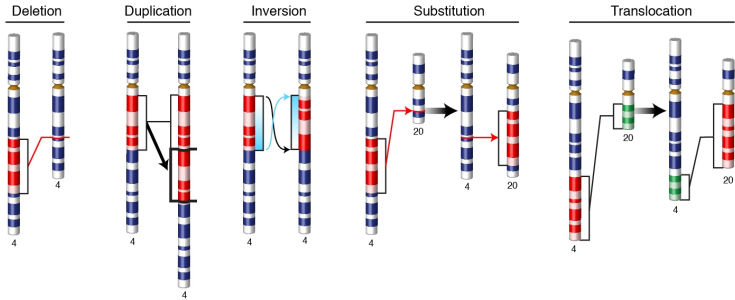
- ① Entwindung der DNA (**Topoisomerasen**)
- ② Auftrennung des DNA-Doppelstrangs (**Helikase**)
- ③ Synthese der RNA-Primer (**Primasen**)
- ④ Kopieren der beiden Elternstränge ausgehend von den RNA-Primern (**DNA-Polymerasen**)
 - ⇒ es entstehen zwei komplementäre Tochterstränge
 - ⇒ kontinuierliche Synthese des Leitstrangs
 - ⇒ diskontinuierliche Synthese des Folgestrangs (Okazaki-Fragmente)

DNA-Replikation

Natürlicher Vorgang zur Vervielfältigung der DNA bei der Zellteilung:

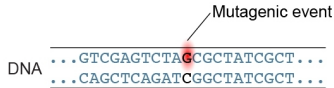
- ① Entwindung der DNA (**Topoisomerasen**)
- ② Auftrennung des DNA-Doppelstrangs (**Helikase**)
- ③ Synthese der RNA-Primer (**Primasen**)
- ④ Kopieren der beiden Elternstränge ausgehend von den RNA-Primern (**DNA-Polymerasen**)
 - ⇒ es entstehen zwei komplementäre Tochterstränge
 - ⇒ kontinuierliche Synthese des Leitstrangs
 - ⇒ diskontinuierliche Synthese des Folgestrangs (Okazaki-Fragmente)
- ⑤ Verbindung der Okazaki-Fragmente des Folgestrangs (**Ligase**)

Mutationen



Bildquelle: [6]

Mutationen



Deletion



Insertion

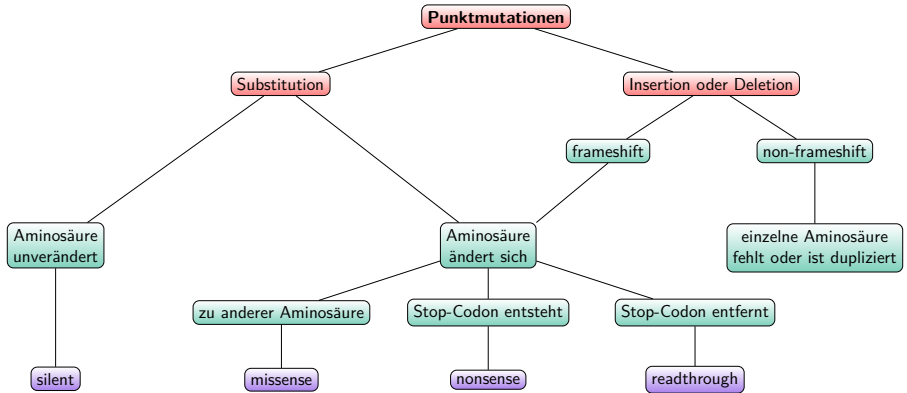


Substitution



Bildquelle: [6]

Mutationen



Folgen von Mutationen

- Loss-of-function-Mutationen
- Gain-of-function-Mutationen

Folgen von Mutationen

- Loss-of-function-Mutationen
- Gain-of-function-Mutationen

Varianten

- oft Varianten einzelner Basen: SNPs (single nucleotide polymorphism)
- ohne pathologische Auswirkungen
- vermehrtes Auftreten innerhalb einer Spezies

Folgen von Mutationen

- Loss-of-function-Mutationen
- Gain-of-function-Mutationen

Varianten

- oft Varianten einzelner Basen: SNPs (single nucleotide polymorphism)
- ohne pathologische Auswirkungen
- vermehrtes Auftreten innerhalb einer Spezies

Allele, Locus, Ploidie und Genotyp

- **Allele:** verschiedene Varianten eines genomischen Ortes (**Locus**)
- **Ploidie:** Anzahl der Chromosomensätze (**homologe Chromosomen**)
- **Genotyp:**
 - ⇒ **Homozygotie:** an einem Locus liegt auf allen homologen Chromosomen das gleiche Allel vor
 - ⇒ **Heterozygotie:** die homologen Chromosomen weisen an einem Locus unterschiedliche Allele auf

Polymerase-Kettenreaktion (PCR)

- Methode zur Vervielfältigung von DNA-Abschnitten
- mehrere Zyklen der folgenden temperaturabhängigen Schritte:

Polymerase-Kettenreaktion (PCR)

- Methode zur Vervielfältigung von DNA-Abschnitten
- mehrere Zyklen der folgenden temperaturabhängigen Schritte:
 - ① **Denaturierung:** durch Erhitzen wird der DNA-Doppelstrang in zwei Einzelstränge aufgespalten (96°C)

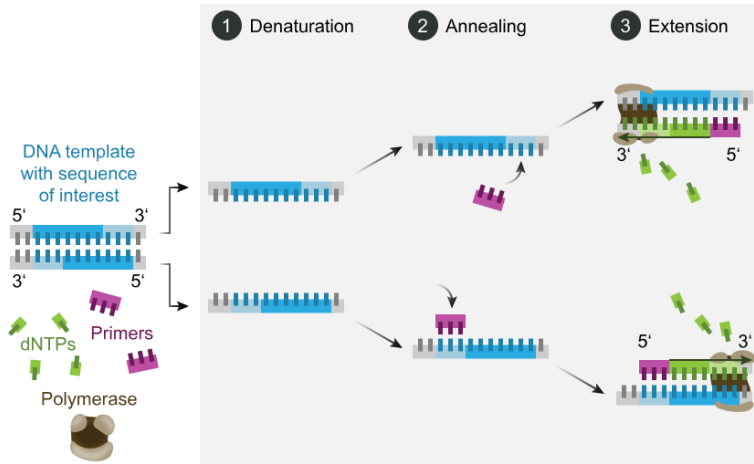
Polymerase-Kettenreaktion (PCR)

- Methode zur Vervielfältigung von DNA-Abschnitten
- mehrere Zyklen der folgenden temperaturabhängigen Schritte:
 - 1 **Denaturierung:** durch Erhitzen wird der DNA-Doppelstrang in zwei Einzelstränge aufgespalten (96°C)
 - 2 **Annealing:** Primerbindung an den 3'-Enden der zu amplifizierenden Gensequenz beider Einzelstränge (55-65°C)

Polymerase-Kettenreaktion (PCR)

- Methode zur Vervielfältigung von DNA-Abschnitten
- mehrere Zyklen der folgenden temperaturabhängigen Schritte:
 - 1 **Denaturierung:** durch Erhitzen wird der DNA-Doppelstrang in zwei Einzelstränge aufgespalten (96°C)
 - 2 **Annealing:** Primerbindung an den 3'-Enden der zu amplifizierenden Gensequenz beider Einzelstränge (55-65°C)
 - 3 **Elongation:** DNA-Synthese der komplementären Stränge (72°C)

Polymerase-Kettenreaktion (PCR)



Bildquelle: [7]

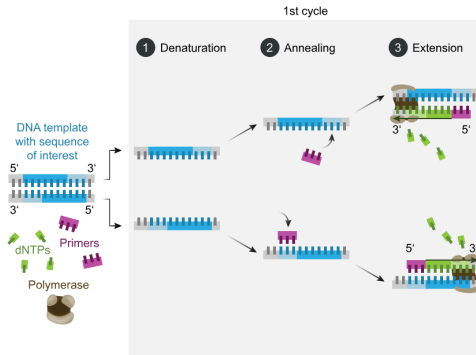
Polymerase-Kettenreaktion (PCR)

- Methode zur Vervielfältigung von DNA-Abschnitten
- mehrere Zyklen der folgenden temperaturabhängigen Schritte:
 - 1 **Denaturierung:** durch Erhitzen wird der DNA-Doppelstrang in zwei Einzelstränge aufgespalten (96°C)
 - 2 **Annealing:** Primerbindung an den 3'-Enden der zu amplifizierenden Gensequenz beider Einzelstränge (55-65°C)
 - 3 **Elongation:** DNA-Synthese der komplementären Stränge (72°C)

Polymerase-Kettenreaktion (PCR)

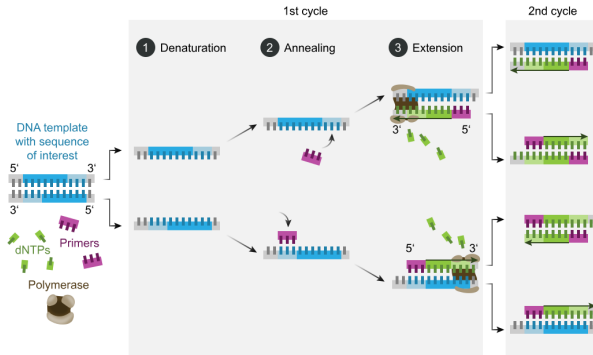
- Methode zur Vervielfältigung von DNA-Abschnitten
- mehrere Zyklen der folgenden temperaturabhängigen Schritte:
 - 1 **Denaturierung**: durch Erhitzen wird der DNA-Doppelstrang in zwei Einzelstränge aufgespalten (96°C)
 - 2 **Annealing**: Primerbindung an den 3'-Enden der zu amplifizierenden Gensequenz beider Einzelstränge (55-65°C)
 - 3 **Elongation**: DNA-Synthese der komplementären Stränge (72°C)
- mit jedem Zyklus wird die betreffende Sequenz verdoppelt
- in Abhängigkeit von der Anzahl der durchgeführten Zyklen n exponentieller Anstieg der Kopien 2^n

Polymerase-Kettenreaktion (PCR)



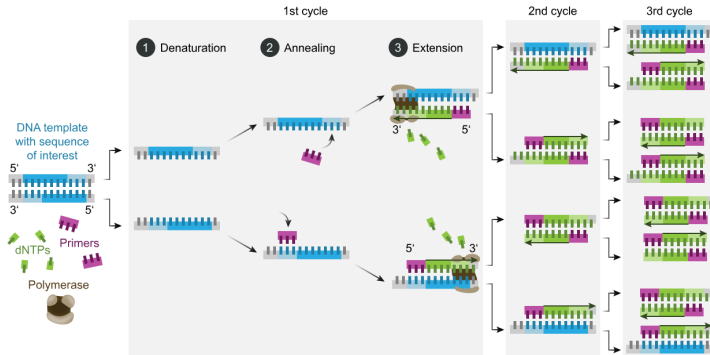
Bildquelle: [7]

Polymerase-Kettenreaktion (PCR)



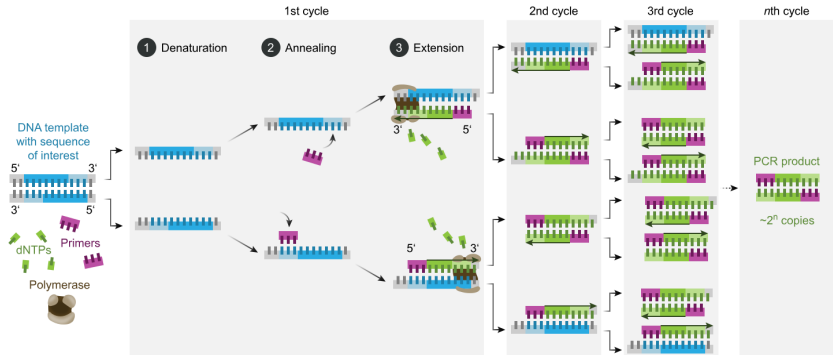
Bildquelle: [7]

Polymerase-Kettenreaktion (PCR)



Bildquelle: [7]

Polymerase-Kettenreaktion (PCR)



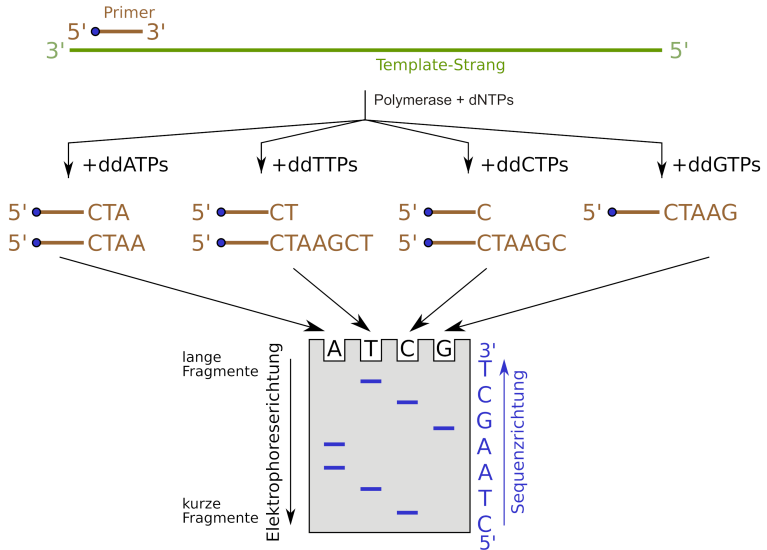
Bildquelle: [7]

Sequenzierung

Sanger-Sequenzierung

Kettenabbruch-Synthese mit vier Probenansätzen denen jeweils eine der vier möglichen Nukleotide in modifizierter Form beigefügt wird

Sequenzierung



Bildquelle: [8]

Sequenzierung

Sanger-Sequenzierung

Kettenabbruch-Synthese mit vier Probenansätzen denen jeweils eine der vier möglichen Nukleotide in modifizierter Form beigefügt wird

Sequenzierung

Sanger-Sequenzierung

Kettenabbruch-Synthese mit vier Probenansätzen denen jeweils eine der vier möglichen Nukleotide in modifizierter Form beigefügt wird

NGS-Sequenzierung

verbesserte Sequenziertechnologien im Hochdurchsatzverfahren

Sequenzierung

Sanger-Sequenzierung

Kettenabbruch-Synthese mit vier Probenansätzen denen jeweils eine der vier möglichen Nukleotide in modifizierter Form beigefügt wird

NGS-Sequenzierung

verbesserte Sequenziertechnologien im Hochdurchsatzverfahren

RAD-Sequenzierung

- restriction site associated DNA sequencing

Sequenzierung

Sanger-Sequenzierung

Kettenabbruch-Synthese mit vier Probenansätzen denen jeweils eine der vier möglichen Nukleotide in modifizierter Form beigefügt wird

NGS-Sequenzierung

verbesserte Sequenziertechnologien im Hochdurchsatzverfahren

RAD-Sequenzierung

- restriction site associated DNA sequencing
- **Anwendung:** Populationsgenetik, Ökologie, Genotypisierung, Evolutionsforschung

Sequenzierung

Sanger-Sequenzierung

Kettenabbruch-Synthese mit vier Probenansätzen denen jeweils eine der vier möglichen Nukleotide in modifizierter Form beigelegt wird

NGS-Sequenzierung

verbesserte Sequenziertechnologien im Hochdurchsatzverfahren

RAD-Sequenzierung

- restriction site associated DNA sequencing
- **Anwendung:** Populationsgenetik, Ökologie, Genotypisierung, Evolutionsforschung
- Sequenzierung multipler kleiner DNA-Fragmente aus dem gesamten Genom, ohne vollständige genomische Abdeckung

Sequenzierung

Sanger-Sequenzierung

Kettenabbruch-Synthese mit vier Probenansätzen denen jeweils eine der vier möglichen Nukleotide in modifizierter Form beigelegt wird

NGS-Sequenzierung

verbesserte Sequenziertechnologien im Hochdurchsatzverfahren

RAD-Sequenzierung

- restriction site associated DNA sequencing
- **Anwendung:** Populationsgenetik, Ökologie, Genotypisierung, Evolutionsforschung
- Sequenzierung multipler kleiner DNA-Fragmente aus dem gesamten Genom, ohne vollständige genomische Abdeckung
- gleichzeitige Analyse mehrerer Individuen in gepoolten Proben

Sequenzierung

Sanger-Sequenzierung

Kettenabbruch-Synthese mit vier Probenansätzen denen jeweils eine der vier möglichen Nukleotide in modifizierter Form beigelegt wird

NGS-Sequenzierung

verbesserte Sequenzieretechnologien im Hochdurchsatzverfahren

RAD-Sequenzierung

- restriction site associated DNA sequencing
- **Anwendung:** Populationsgenetik, Ökologie, Genotypisierung, Evolutionsforschung
- Sequenzierung multipler kleiner DNA-Fragmente aus dem gesamten Genom, ohne vollständige genomische Abdeckung
- gleichzeitige Analyse mehrerer Individuen in gepoolten Proben
- benötigt kein Referenzgenom

Restriktionsenzyme

- molekulare Scheren, welche die DNA an spezifischen Sequenzen schneiden

Restriktionsenzyme

- molekulare Scheren, welche die DNA an spezifischen Sequenzen schneiden
- Enden können glatt oder versetzt sein

Restriktionsenzyme

- molekulare Scheren, welche die DNA an spezifischen Sequenzen schneiden
- Enden können glatt oder versetzt sein

CCCCGGG
GGGCCC

Bildquelle: [8]

Restriktionsenzyme

- molekulare Scheren, welche die DNA an spezifischen Sequenzen schneiden
- Enden können glatt oder versetzt sein

GAATTC
CTTAAG

Bildquelle: [8]

RAD-Verfahren

Restriktionsenzyme

- molekulare Scheren, welche die DNA an spezifischen Sequenzen schneiden
- Enden können glatt oder versetzt sein

Methode

- 1 **DNA-Verdau** durch Restriktionsenzyme

RAD-Verfahren

Restriktionsenzyme

- molekulare Scheren, welche die DNA an spezifischen Sequenzen schneiden
- Enden können glatt oder versetzt sein

Methode

- 1 **DNA-Verdau** durch Restriktionsenzyme
- 2 Sequenz der Restriktionsstelle ist bekannt, dies ermöglicht die Bindung der **Adapter-** und **Barcode**sequenzen

RAD-Verfahren

Restriktionsenzyme

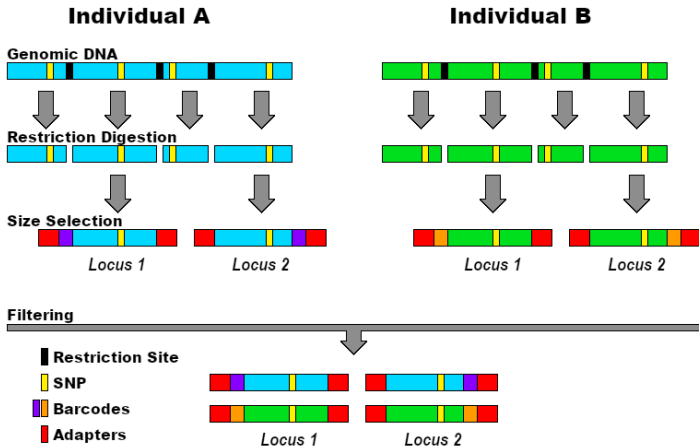
- molekulare Scheren, welche die DNA an spezifischen Sequenzen schneiden
- Enden können glatt oder versetzt sein

Methode

- 1 **DNA-Verdau** durch Restriktionsenzyme
- 2 Sequenz der Restriktionsstelle ist bekannt, dies ermöglicht die Bindung der **Adapter-** und **Barcode**sequenzen
- 3 sgdf

RAD-Verfahren

Restriction-site Associate DNA Sequencing (RADSeq)



Bildquelle: [9] (modifiziert)

Problemstellung und Lösungsansatz

text

text

text

text

text

text

text

text

text

text

text

text

text

text

text

text

text

text

text

text

text

text

text

text

text

text

text

text

text

text

text

text

text

text

text

text

text

text

text

text

Bildquellen I

- [1] MOUAGIP: *Aminoacids table.svg*. 2021. – source: https://commons.wikimedia.org/wiki/File:Aminoacids_table.svg
- [2] MARGULIES, Elliott: *Transcription*. – source: <https://www.genome.gov/genetics-glossary/Transcription>
- [3] LEJA, Darryl: *Transfer RNA (tRNA)*. – source: <https://medlineplus.gov/genetics/understanding/basics/noncodingdna/>
- [4] MARGULIES, Elliott: *Transfer RNA (tRNA)*. – source: <https://www.genome.gov/genetics-glossary/Transfer-RNA>
- [5] RUIZ, Mariana: *DNA replication*. – source: https://commons.wikimedia.org/wiki/File:DNA_replication_en.svg
- [6] COLLINS, Francis: *Mutation*. – source: <https://www.genome.gov/genetics-glossary/Mutation>

- [7] ENZOKLOP: *Polymerase Chain Reaction - Schematic mechanism of PCR*. – source: https://en.wikipedia.org/wiki/File:Polymerase_chain_reaction-en.svg
- [8] CHRISTOPH GOEMANS, Norman M.: *Prinzip der DNA-Sequenzierung nach der Didesoxy-Methode*. – source: <https://de.wikipedia.org/wiki/Datei:Didesoxy-Methode.svg>
- [9] CLARK, Jonathan: *Schematic diagram of RADseq*. – source: https://en.wikipedia.org/wiki/File:RADseq_schematic.pdf