

**Univerzita Karlova**

**Přírodovědecká fakulta**

Studijní obor: Biologie

Studijní program: Molekulární biologie a biochemie organismů



**Sára Buryšová**

Současné genomické a cytogenomické metody v analýze  
chromozómových přestaveb

Current genomic and cytogenomic methods in the analysis  
of chromosome rearrangements

Bakalářská práce

Školitel: **Mgr. Alexandr Sember, Ph. D.**

Praha, 2022/23

## **Poděkování**

Na tomto místě bych v první řadě ráda poděkovala svému školiteli, Mgr. Alexandru Semberovi, Ph. D., který mě provedl celým procesem psaní bakalářské práce a byl vždy ochotný podat mi pomocnou ruku a ukázat správný směr. Další dík patří také Mgr. Marianu Novotnému, Ph. D. za jeho cenné připomínky k bioinformatické části této práce. V neposlední řadě bych pak chtěla poděkovat celé své rodině a blízkým, kteří mi byli po celý rok velkou oporou.

## **Prohlášení**

Prohlašuji, že jsem závěrečnou práci zpracovala samostatně a že jsem uvedla všechny použité informační zdroje a literaturu. Tato práce ani její část nebyla předložena k získání jiného nebo stejného akademického titulu.

V Praze, dne 3. 8. 2023

Podpis:

## **Abstrakt**

Chromozómové přestavby jsou jedním z hlavních mechanismů evoluce eukaryotických genomů a mohou významně přispět k reprodukční izolaci a diverzifikaci, včetně evoluce komplexních znaků přispívajících např. k lokální adaptaci. Jsou to strukturní změny vedoucí ke změně morfologie a/nebo počtu chromozómů, které mohou mít přímý vliv na evoluci a expresní profily genů, četnost a distribuci rekombinace v genomu a funkční dynamiku procesů v interfázním jádře. Mohou být ovšem také podstatou (nebo doprovodným jevem) vrozených vad a nádorového bujení. Studium chromozómových přestaveb a mechanismů jejich vzniku souvisí s identifikací a charakterizací oblastí, kde došlo ke dvouvláknovému zlomu a opětovnému spojení částí chromozómů. Jednou z možností je analyzovat chromozómy a mezidruhové změny v uspořádání vazebných skupin/syntenních bloků pomocí cytogenetických a cytogenomických metod (např. mezidruhovým mapováním celochromozómových hybridizačních sond). Detailnější vhled přináší srovnávací genomika, v dnešní době zastoupená především tzv. metodami třetí generace (platformy firem PacBio a Oxford Nanopore Sequencing), často v kombinaci s dalšími pokročilými metodami jako jsou Hi-C nebo optické mapování. Cílem této bakalářské práce je shrnout současné přístupy v analýze chromozómových přestaveb (včetně strukturních variant), jejich genetického obsahu a míry zachování konzervované syntenie mezi druhy, srovnat výhody a nevýhody jednotlivých přístupů a zhodnotit komplementaritu mezi cytogenetickými a genomickými metodami.

### **Klíčová slova:**

chromozómová fúze, genetické mapování, evoluční chromozómové zlomy, Hi-C, inverze, konzervovaná syntenie, optické mapování, sekvenování třetí generace, strukturní varianty, Zoo-FISH

## **Abstract**

Chromosome rearrangements represent one of the major mechanisms driving the eukaryotic genome evolution. They may significantly contribute to reproductive isolation and diversification, including the evolution of complex life-history traits linked e.g. with local adaptation. They are structural changes leading to alteration in the morphology and/or number of chromosomes, which can have a direct effect on the evolution of genes and their expression profiles, the frequency and distribution of recombination in the genome, and the functional dynamics of processes operating in the interphase nucleus. However, they may be also causal for etiology (or a consequence) of inherent diseases and tumorigenesis. The study of chromosome rearrangements and the mechanisms of their emergence is related to the identification and characterization of rearrangement breakpoints (i.e. where double-strand break occurred and chromosome segments subsequently rejoined). One possibility is to analyze chromosomes and interspecific changes in the arrangement of linkage groups/synteny blocks using cytogenetic and cytogenomic methods (e.g. cross-species mapping of whole-chromosome hybridization probes). More detailed insight is provided by comparative genomics, nowadays mainly represented by so-called third-generation methods (the platforms of PacBio and Oxford Nanopore Sequencing), often in combination with other advanced methods such as Hi-C or optical mapping. The aim of this bachelor thesis is to summarize current approaches in the analysis of chromosome rearrangements (including structural variants), their genetic content and the degree of conserved synteny between species, to compare the advantages and disadvantages of individual approaches and to evaluate the complementarity between cytogenetic and genomic methods.

### **Key words:**

chromosome fusion, genetic mapping, evolutionary breakpoints, Hi-C, inversion, conserved synteny, optical mapping, third-generation sequencing, structural variants, Zoo-FISH

## Seznam použitých zkratek

- BAC** – umělý bakteriální chromozóm, z angl. *bacterial artificial chromosome*
- BIR** – zlom indukující replikaci, z angl. *break-induced replication*
- BUSCO** – z angl. *Benchmarking Universal Single-Copy Orthologs*
- CCS** – kruhové konsenzuální čtení, z angl. *circular consensus reads*
- CLR** – souvislé dlouhé čtení, z angl. *continuous long reads*
- CNVs** – změny počtu kopií, z angl. *copy number variations*
- CO** – z angl. *crossing over*
- CRs** – chromozómové přestavby, z angl. *chromosomal rearrangements*
- CSA** – z angl. *Chromosome-Scale Assembler*
- CsgG** – *Curlin sigma S-dependent grow*
- DLE** – z angl. *Direct Label Enzyme*
- DLS** – z angl. *Direct Label and Stain*
- dNTP** – deoxyribonukleosidtrifosfát, z angl. *deoxynucleoside triphosphate*
- DSB** – dvouřetězcový zlom, z angl. *double stranded break*
- EBRs** – evolučně konzervované dvouvláknové zlomy, z angl. *evolutionary breakpoint regions*
- FISH** – fluorescenční hybridizace *in situ*, z angl. *fluorescence in situ hybridization*
- FoSTeS** – zastavení replikační vidličky a výměna templátů, z angl. *fork stalling and template switching*
- GATK** – sada nástrojů pro analýzu genomu, z angl. *genome analysis toolkit*
- HiFi** – vysoká přesnost, z angl. *High Fidelity*
- Hi-C** – vysoce výkonné zachycení chromozómové konformace, z angl. *high throughput chromosome conformation capture*
- HSBs** – homologní syntenní bloky, z angl. *homologous synteny blocks*
- NAHR** – nealelická homologní rekombinace, z angl. *non-allelic homologous recombination*
- NF** – fundamentální číslo, z fr. *Nombre Fondamental*
- NGS** – sekvenování druhé generace, z angl. *Next generation sequencing*
- NHEJ** – nehomologní spojení konců, z angl. *Non-homologous end joining*
- ONT** – nanopórové technologie z Oxfordu, z angl. *Oxford Nanopore Technologies*
- OLC** – z angl. *Overlapping layout consensus*
- PCR** – polymerázová řetězcová reakce, z angl. *polymerase chain reaction*
- RD** – hloubka čtení, z angl. *read depth*
- RP** – páry čtení, z angl. *read pair*
- SNP** – polymorfismus na úrovni jednoho nukleotidu, z angl. *single nucleotide polymorphism*
- SMRT** – sekvenování jedné molekuly v reálném čase, z angl. *single molecule realtime sequencing*
- SR** – rozdělené čtení, z angl. *split reads*
- SSA** – spojování jednořetězcového vlákna, z angl. *single strand annealing*
- ssDNA** – jednořetězcová molekula DNA, z angl. *single-strand DNA*
- SVs** – strukturní varianty, z angl. *structural variants*

**TADs** – topologicky asociované domény, z angl. *topologically associating domains*

**TGS** – sekvenování třetí generace, z angl. *Third generation sequencing*

**ULR** – ultra dlouhá čtení, z angl. *ultra long reads*

**VGP** – projekt genomu obratlovců, z angl. *Vertebrate genome project*

**ZMW** – z angl. *zero mode waveguide*

# Obsah

1	Úvod.....	1
1.1	Cíle bakalářské práce.....	1
2	Chromozómové přestavby.....	2
2.1	Delece a duplikace.....	2
2.2	Inverze.....	4
2.3	Inzerce.....	5
2.4	Translokace.....	5
2.5	Fúze a centrické rozpady.....	5
2.6	Polymorfizmy a mechanizmy fixace chromozómových přestaveb.....	8
2.7	Strukturní varianty.....	9
2.8	Chromozómové přestavby a jejich úloha v evoluci.....	9
2.8.1	Vliv CRs na funkci a evoluci genů.....	11
2.9	EBRs a fragilní místa.....	12
3	Metody zkoumání CRs.....	12
3.1	Cytogenetika a komparativní cytogenomika.....	13
3.2	Genomické metody.....	14
3.2.1	Sekvenování druhé generace.....	14
3.2.2	Sekvenování třetí generace.....	17
3.2.3	Metoda Hi-C.....	21
3.2.4	Optické mapování.....	23
3.3	Sestavování genomu.....	24
3.3.1	Rekonstrukce genomu společných předků.....	25
3.4	Algoritmy pro detekci SVs.....	26
3.4.1	Alligners.....	26
3.4.2	Callers.....	27
3.5	Bioinformatické pipeline.....	28
4	Závěr.....	29
	Literatura.....	32
	Internetové zdroje.....	41

# 1 Úvod

Každý eukaryotický druh má svoji genetickou informaci v jádře buňky organizovanou do specifického, mnohdy jedinečného počtu a morfologie chromozómů. V průběhu evoluce se od sebe příbuzné druhy v tomto ohledu postupně odlišují, mimo jiné vlivem chromozómových přestaveb (neboli CRs, z angl. *chromosome rearrangements*). CRs jsou změny ve struktuře chromozómů, které mohou mít vliv na architekturu jádra z hlediska počtu a prostorové organizace chromozómů, dále na evoluci a expresi genů v okolí přestavby, rekombinačních vzorů a také mohou v konečném důsledku vést k mezidruhové reprodukční izolaci (Navarro & Barton, 2003; Damas et al., 2021). Na úrovni jedince však mohou mít CRs fatální důsledky z hlediska životaschopnosti nebo fertility a dále mohou být také kauzálně spojeny s nádorovou transformací (Wilhelm et al., 2018; Tuna et al., 2019). Klíčové informace o mechanismech vzniku a důsledcích CRs přináší vedle cytogenetických metod především analýza sekvencí DNA v oblastech chromozómů, kde došlo ke dvouvláknovému zlomu (neboli DSB, z angl. *double-stranded break*) a jeho následné chybné opravě. Identifikace a interpretace takových míst však zůstává i nadále velkou bioinformatickou výzvou. V posledních letech došlo k výraznému pokroku ve vývoji vhodných genomických přístupů a asociovaných bioinformatických programů, které umožnily vedle přesnějšího a kompletnějšího poskládání genomů také lépe pochopit principy prostorové a funkční organizace chromozómů v jádře jakožto i mechanismů a důsledků, které jsou spjaty se změnami tohoto uspořádání (Sedlazeck et al., 2018a; Mahmoud et al., 2019; Damas et al., 2021).

## 1.1 Cíle bakalářské práce

Tato práce si klade za cíl shrnout a srovnat přínos metod molekulární cytogenetiky a genomiky v odhalování CRs včetně jejich polymorfizmů neboli strukturních variant, jejich genetického obsahu a také míry zachování konzervované syntenie (tj. uspořádání stejného bloku sekvencí ve stejné vazebné skupině v průběhu evoluce) meziblízce či vzdáleně příbuznými organizmy.

## 2 Chromozómové přestavby

Eukaryotní organizmy se liší nejen velikostí a skladbou sekvencí svých genomů, ale také mírou fragmentace genomu na jednotlivé chromozómy. Napříč druhy existuje obrovská variabilita počtu a morfologie chromozómů, tj. charakteristik souhrnně označovaných jako karyotyp. Počty chromozómů se obvykle vztahují k somatickým buňkám (tj. uvádí se diploidní stav,  $2n$ ). Morfologie chromozómů souvisí s pozicí centromery, tj. chromozómy jednoramenné mají centromeru výrazně blíže jednomu okraji chromozómu (a tedy krátká raménka jsou výrazně kratší než raménka dlouhá) zatímco chromozómy dvouramenné (metacentrické a submetacentrické) mají centromeru situovanou více intersticiálně (Levan et al., 1964). Ke změnám karyotypu dochází postupně v průběhu evoluce a mohou souviset se změnou ploidie (tj. polyploidizace), obsahu repetitivní DNA a s chromozómovými přestavbami (CRs) (Mayrose & Lysák, 2021; Heslop-Harrison et al., 2023). Tyto mechanismy mohou vést k reprodukční izolaci mezi jedinci různých populací stejného druhu/blízce příbuzných druhů (např. vlivem chybného párování a rozchodu chromozómů v meióze, a tedy zablokováním meiózy v kontrolních bodech), ale mohou také podporovat diverzifikaci a adaptační potenciál na dané podmínky prostředí (Navarro & Barton, 2003; Pellestor et al., 2011). Ve své práci bych se chtěla specificky zaměřit na CRs, tj. strukturní změny chromozómů, které jsou důsledkem chybných oprav dvouvláknových zlomů a chyb při replikaci a rekombinaci (Tsai & Lieber, 2010; Weckselblatt & Rudd, 2015). CRs se odehrávají v rámci jednoho (intrachromozómové) nebo více (interchromozómové) elementů v dané buňce. CRs se dále rozlišují na základě toho, zda v důsledku přestavby došlo ke změně množství genetické informace. Při ztrátě (delece) nebo zisku (duplikace, amplifikace) hovoříme o nebalancované přestavbě. Druhou skupinou jsou přestavby balancované (např. translokace, fúze/rozpady či inverze).

Primární příčinou CRs je celá řada. Nejčastěji jsou způsobovány chybnými opravami dvouřetězcových zlomů DNA (např. mechanismy NAHR, nealelická homologní rekombinace, z angl. *non-allelic homologous recombination* nebo NHEJ, nehomologní spojení konců, z angl. *non-homologous end joining*), chybami během replikace, meiotické či mitotické homologní rekombinace, chybou v působení topoizomeráz, vlivem kyslíkových radikálů z mitochondrií nebo vnějších faktorů (nazývaných v tomto textu jako klastogeny) jakou jsou různé chemikálie, záření nebo viry (Schubert & Lysák, 2011; Morin et al., 2017; Poot & Hochstenbach, 2021).

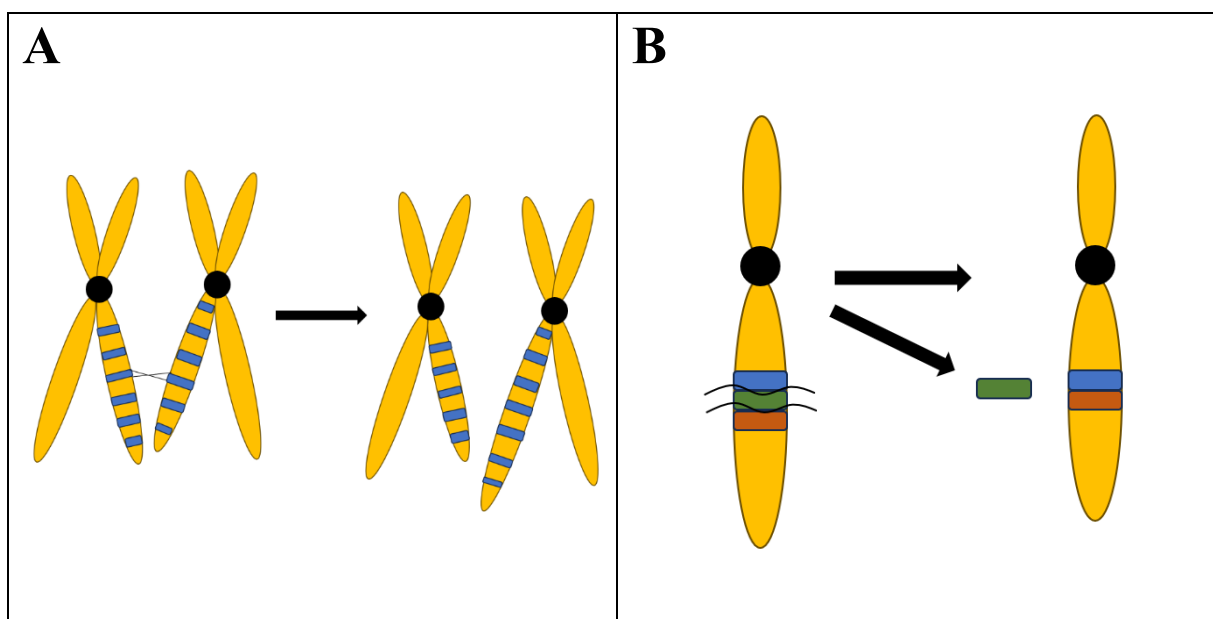
V následujících kapitolách představím základní typy CRs, popíšu stručně mechanismy jejich vzniku a fixace v populaci. Budu se zabývat také rolí CRs v reprodukční izolaci, diverzifikaci a adaptaci. Mezi základní strukturní přestavby patří: delece, duplikace/amplifikace, inverze, inzerce, různé typy translokací a fúzí a centromerické rozpady (Schubert & Lysák, 2011).

### 2.1 Delece a duplikace

Duplikace (případně amplifikace) znamená zmnožení úseku DNA, zatímco delecí se rozumí ztráta úseku chromozómu. Základním mechanismem, který vede současně k delecí na jedné a duplikaci

na druhé kopii chromozómu, je nerovnoměrný crossing-over (CO), který obvykle probíhá mezi kopiemi tandemových repetit, jež k sobě nepřiléhají ve stejném pořadí na rekombinujících molekulách DNA. K nerovnoměrnému CO může docházet jak mezi homologními chromozómy, tak mezi chromatidami stejného chromozómu (tj. výměna sesterských chromatid) (Obr. 1A). Delece a duplikace mohou být dále důsledkem procesů oprav dvouvláknových zlomů, např. z důvodu exonukleolytického štěpení (mechanizmy NHEJ a SSA, z angl. *single-strand annealing*) (Lysák & Schubert, 2013; Anand et al., 2014) nebo přeskočku mezi templátovými molekulami DNA [mechanizmy BIR, z angl. *break-induces replication* (Leffak, 2017) a FoSTeS, z angl. *fork stalling and template switching* (F. Zhang et al., 2009)].

Delece rozlišujeme podle toho, jestli byl ztracen koncový úsek (terminální delece) nebo vnitřní segment chromozómu (intersticiální delece, Obr. 1B). V případě terminální delece je nutné, aby si buňka dosyntetizovala nové telomerické oblasti, jinak může dojít k fúzi s jiným chromozómem nebo ke vzniku kruhového chromozómu (Yu & Graf, 2010). Delece mají svůj specifický význam u polyploidů, kde dochází k eliminaci různých neesenciálních duplikátů genomu (často repetit), a to napomáhá postupné re-diploidizaci genomu (Dodsworth et al., 2016; Mandáková & Lysák, 2018).



**Obrázek 1:** A) Duplikace a delece způsobená nerovnoměrným meiotickým crossing overem mezi homologními chromozómy, B) Intersticiální delece, je ztracen segment z původního chromozómu a dochází ke ztrátě genetického materiálu, upraveno podle Lysák & Schubert (2013).

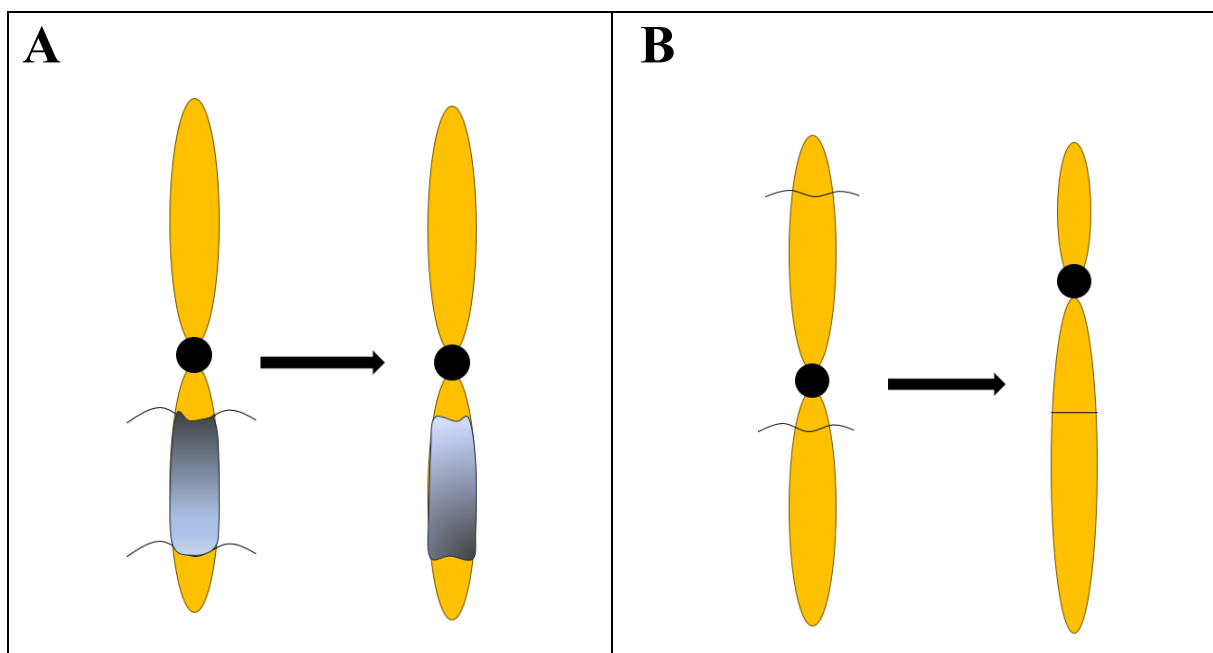
Existují dva typy duplikací, inzerční (neboli ektopická) a tandemová. Inzerční duplikace znamená, že se oddělený zduplikovaný segment přemístí na jiný chromozóm, zatímco při tandemové duplikaci se nová kopie zařadí hned vedle původní sekvence na stejném chromozómu. Specifickým typem je dále segmentální duplikace, která je definovaná jako duplikace oblasti dlouhé minimálně 1 kb, přičemž kopie mají relativně vysokou sekvenční identitu (Abdullaev et al., 2021). Vedle dobře

popsaného vlivu duplikací na evoluci genomu (viz dále) je zřejmé, že duplikace/amplifikace především repetitivních úseků zvyšuje pravděpodobnost dalších přestaveb (King, 1993; S.-F. Li et al., 2017). Navíc přítomnost kopií stejné repetice na různých chromozómech může vést k rekombinaci mezi nimi (tj. ektopická neboli nealelická rekombinace) a výsledkem mohou být různé typy přestaveb jako například translokace, fúze (v rámci dvou chromozómů) nebo delece (v rámci jednoho chromozómu, kdy orientace rekombinujících kopií je stejná, tj. jsou to tzv. „přímé“ repetice) nebo inverze (v rámci chromozómu rekombinují kopie repetice, jejichž sekvence jsou vůči sobě v opačné orientaci, tzv. invertované repetice) (Delprat et al., 2009; S.-F. Li et al., 2017; Balachandran et al., 2022).

## 2.2 Inverze

Inverze je jednou z nejčastěji popisovaných CRs, jedná se dokonce o první historicky popsanou strukturní přestavbu (Sturtevant, 1917). Inverzi můžeme popsat jako přestavbu, při které dojde ke dvěma zlomům DNA v rámci jednoho chromozómu, otočení prostředního segmentu o  $180^\circ$  a jeho následnému opětovnému spojení se zbývajícím úseky chromozómu. Alternativně může dojít k ektopické rekombinaci (viz výše). Segment, který prodělal inverzi, je nyní v opačné orientaci. V konečném důsledku se tedy mění pořadí genů na invertovaném úseku.

Inverze lze rozlišovat podle toho, zda invertovaný úsek zahrnuje centromeru. U paracentrické inverze (Obr. 2A) ji nezahrnuje, zatímco v případě pericentrické ano (Obr. 2B). V konečném důsledku, zatímco paracentrická inverze nemění morfologii, pericentrická inverze může měnit poměr délky ramének, pokud oba zlomy DNA mají různou vzdálenost od centromery (Lysák & Schubert, 2013). Je ale již dobře známo, že pozice centromery se může měnit i bez inverze, kdy původní centromera zanikne a její roli převezme jiný úsek chromozómu (tzv. centromerický posun) (Schubert, 2018).



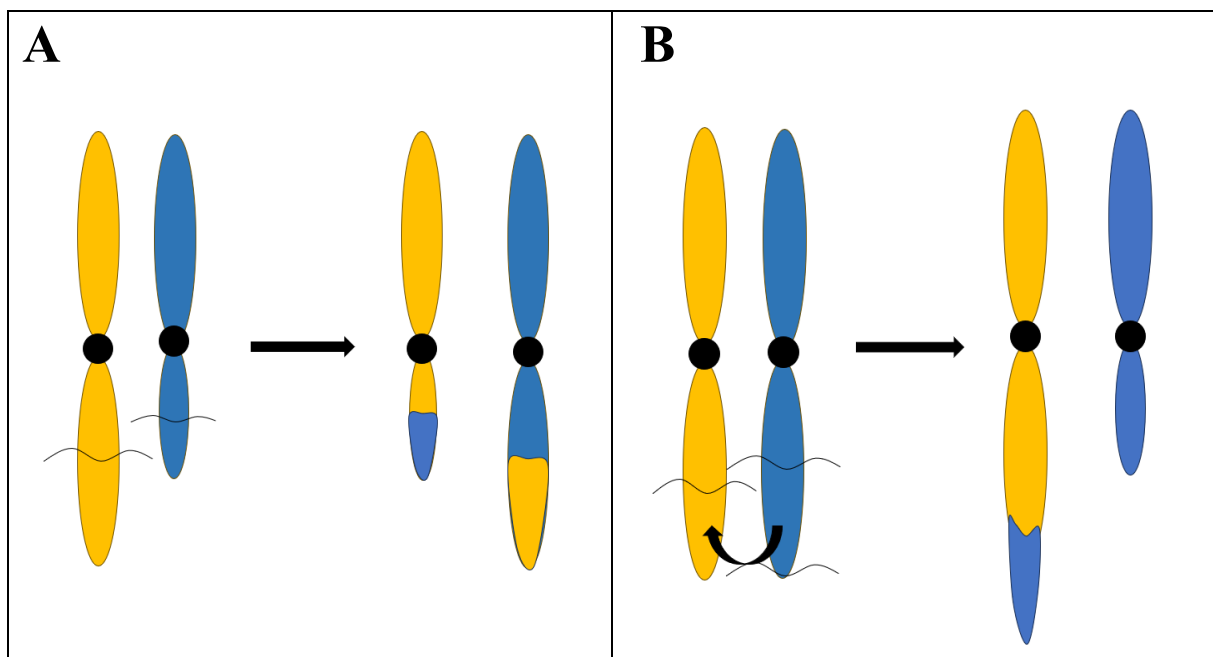
**Obrázek 2:** A) Paracentrická inverze, B) Pericentrická inverze, upraveno podle (Lysák & Schubert, 2013).

## 2.3 Inzerce

Inzerce, jinak také nazývaná inzerční translokace, je poměrně vzácná interchromozómová přestavba. K inzerci dochází translokací segmentu z jednoho chromozómu a jeho následnému začlenění na nehomologní chromozóm (většinou v intersticiální oblasti) (Weckselblatt & Rudd, 2015). Mechanismus inzerce není zatím podrobně prozkoumán, základním předpokladem pro její průběh je však vytvoření alespoň tří DSBs (Neil et al., 2011).

## 2.4 Translokace

Jako translokace se označují takové přestavby, při kterých dochází k přemístění chromozómových segmentů nebo celých ramen, obvykle mezi dvěma nehomologními chromozómy. Pakliže si oba chromozómy navzájem vymění určitý úsek, jedná se o translokaci reciprokovou (Obr. 3A). Pokud se jednosměrně přesune segment z jednoho chromozómu na druhý, jde o translokaci nerecipkovou (Obr. 3B). Chromozómy s přestavbou a jejich nepozměněné homology (tj. dohromady čtyři elementy) spolu z důvodu sdílených homologních úseků párují v meióze I za vzniku tzv. kvadrivalentu. Z tohoto útvaru se následně chromozómy rozcházejí do dceřiných buněk a mohou takto vzniknout geneticky balancované či nebalancované gamety (Schubert & Lysák, 2011).



Obrázek 3: A) Reciproká translokace, B) Nerecipková translokace, upraveno podle Schubert & Lysák (2011).

## 2.5 Fúze a centrické rozpady

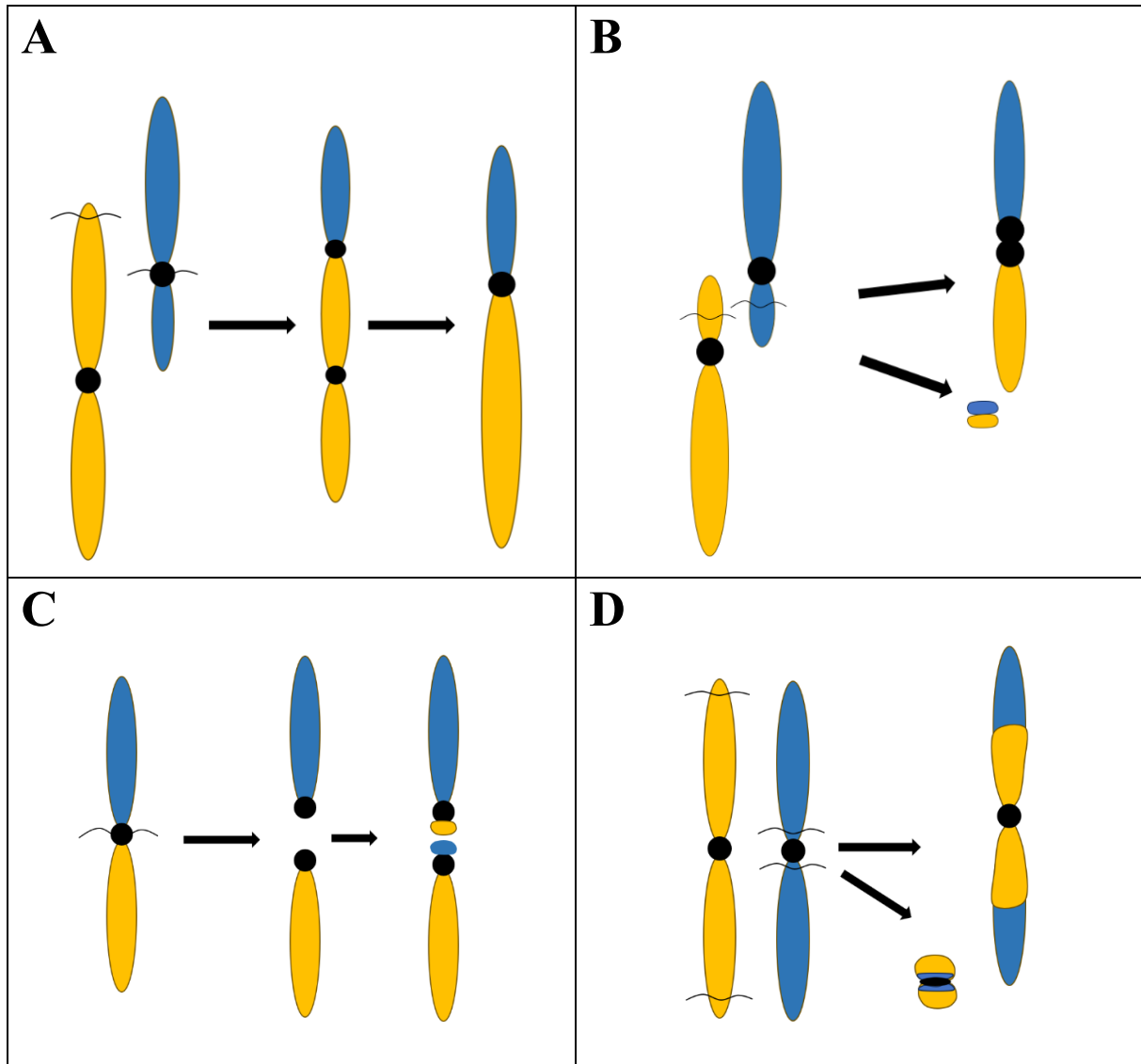
Chromozómy spolu mohou za určitých aberantních okolností fúzovat anebo se rozpadnout v oblasti centromery. Rozpady mimo oblast centromery obvykle znamenají v podstatě terminální delecii, protože fragment bez centromery (tj. acentrický fragment) se s vysokou pravděpodobností vyřadí

v následných buněčných děleních (pokud se na něm nestihne vytvořit neocentromera) (McKinley & Cheeseman, 2016). Chromozómová fúze na druhou stranu reprezentuje svým způsobem extrémní příklad translokace, při které je přemíst'ován celý chromozóm (Schubert & Lysák, 2011).

Rozlišujeme fúze tandemové, centrické a tzv. vmezežené (z angl. „nested“) (King, 1993; Schubert & Lysák, 2011; Lysák, 2022). Při tandemové neboli také end-to-end fúzi se spojují chromozómy, u nichž došlo i) ke zlomu v koncových oblastech chromozómu (nebo k dysfunkci telomery) nebo ii) ke zlomu v centromere jednoho a koncové oblasti druhého chromozómu (Obr. 4A). Nejčastěji se jedná o propojení mezi dvěma jednoramennými chromozómy, nebo mezi jednoramenným a dvouramenným chromozómem. Oproti tomu centrická fúze, která představuje jeden z nejběžnějších typů přestaveb v evoluci především živočichů (viz níže), znamená buď i) přímou fúzi (pomocí opravných mechanismů) mezi jednoramennými chromozómy zbavených funkčních telomer (tj. telomery jsou příliš krátké na tvorbu stabilizačních struktur nebo byly dvouvláknovým zlomem odstraněny) (Slijepcevic, 1998) nebo ii) ektopickou rekombinací mezi (peri)centrickými oblastmi (Schubert & Lysák, 2011; Jarmuz-Szymczak et al., 2014; Guarracino et al., 2023). V konečném důsledku vzniká nápadně velký dvouramenný chromozóm a jeho stabilita v následných buněčných děleních závisí na tom, jestli jsou dvě původní centromery natolik blízko, že vytvoří jeden funkční celek, anebo se jedna z nich inaktivuje [rekombinací nebo epigenetickou inaktivací, např. metylací DNA či histonů (Zhang et al., 2010; Sato et al., 2012; Lysák, 2022)]. Pokud by totiž obě původní centromery zůstaly aktivní a fungovaly nezávisle, tak by hrozilo nebezpečí, že se k oběma připojí mikrotubuly dělicího vřeténka z opačného pólu a budou se následně v průběhu anafáze o chromozóm přetahovat, až se na náhodném místě chromozóm „roztrhne“ a poskytuje nový substrát pro další přestavby, obvykle fúze (tzv. *break-fusion-break* cyklus). Takováto situace může s větší pravděpodobností nastat u tandemových/end-to-end fúzí (přehledy viz McKinley & Cheeseman, 2016; Barra & Fachinetti, 2018; Lysák, 2022), což by mohlo částečně vysvětlovat, proč nejsou v evoluci tak frekventovaně fixované jako centrické fúze (King, 1993; Schmid et al., 2010).

Centrické fúze poprvé pozorované u rovnokřídleho hmyzu W. R. Robertsonem (Robertson, 1916, odtud také pojmenovány Robertsonovy translokace, zkráceně Rb fúze), patří mezi nejčastější přestavby fixované v evoluci eukaryot, především u živočichů (King, 1993; Pardo-Manuel de Villena & Sapienza, 2001; Ferguson-Smith & Trifonov, 2007; Schmid et al., 2010). Jak jsem již zmínila, při této přestavbě dochází ke spojení akrocentrických nebo subtelomerických chromozómů v oblasti centromery za vzniku nápadně velkého metacentrického nebo submetacentrického chromozómu (Obr. 4B). Jedná se tímto vlastně o specifický případ reciproké translokace. Přesný mechanismus závisí na tom, v jakých oblastech došlo na fúzujících chromozómech ke dvouvláknovým zlomům (tj. jestli v rámci centromerických satelitů nebo uvnitř krátkých ramének). Obvykle následně dochází ke ztrátě krátkých ramének, které se v podobě drobného vedlejšího produktu přestavby vyřadí v následných buněčných děleních. Pokud krátká raménka obsahují pouze repetitivní sekvence, tak jejich ztráta nemá zásadní vliv na životaschopnost organismu (Bandyopadhyay et al., 2002; Schubert & Lysák, 2011). Negativní efekt

Rb fúze (nebo koneckonců jakékoliv fúze) souvisí s tím, že se u heterozygotů pro přestavbu v profázi prvního meiotického dělení tvoří trivalent (tj. produkt fúze páruje se dvěma nezfúzovanými homology) a segregace chromozómů do gamet může vést k nebalancovanému stavu (Schmid et al., 2010; Poot & Hochstenbach, 2021).



**Obrázek 4:** A) Tandemová fúze metacentrických a subtelomerického chromozómu a následná inaktivace jedné z centromer fúzního produktu, B) Centrická fúze/Robertsonova translokace při které vzniká metacentrický chromozóm a fragment bez centromery, C) Centrický rozpad na dva chromozómy, které jsou dále stabilizovány adicí telomerických repetitiv, (např. činností telomerázy), D) Vmezežená „nested“ fúze, upraveno podle Lysák & Schubert (2013).

Centrické fúze se dají v karyotypu snadno identifikovat, jelikož jejich vlivem dochází ke snížení  $2n$  za současného navýšení počtu dvouramenných chromozómů, a tato nepřímá korelace je patrná, když se srovnají karyotypy blízké příbuzných druhů. Centrické fúze současně nemění celkový počet chromozómových ramen v karyotypu, takže srovnávání tohoto parametru [pojmenované *Nombre Fondamental*, NF neboli také fundamentální číslo (Matthey, 1945)] a  $2n$  umožňuje kvantifikaci

centrických fúzí. Bez známých fylogenetických vztahů ovšem není možné určit, zda původním stavem byly jednoramenné chromozómy, které zfúzovaly nebo tomu bylo naopak (tj. centrický rozpad dvouramenného chromozómu) (Dobigny et al., 2004). Centromerické rozpady (fize, Obr. 4C) jsou dokumentovány podstatně méně, patrně proto, že produkty této přestavby nejsou v karyotypu tolik nápadné. Navíc oproti fúzím je potřeba, aby oba produkty fize obnovily na svých krátkých raménkách telomery, jinak se jedná o nestabilní produkt, který může podléhat dalším přestavbám (Perry et al., 2004).

Posledním typem fúze je tzv. „nested“ neboli vmezeřená fúze (Obr. 4D). S tímto termínem byl svět poprvé seznámen roku 2009, kdy byla tato přestavba pozorována u krytosemenných rostlin *Poaceae* (Luo et al., 2009). Princip přestavby spočívá v kombinaci dvou nehomologních chromozómů, kdy se jeden celý chromozóm vmezeří do centromerické oblasti druhého chromozómu. Výsledkem je nový di- nebo dokonce tricentrický chromozóm, který je okamžitě stabilizován epigenetickou inaktivací centromery recipientního chromozómu. Aktivní tedy zůstává centromera vloženého chromozómu (Lysák, 2022). Základním předpokladem pro průběh vmezeřené fúze je vznik minimálně tří DSBs, které jsou místně specifické. Dva z těchto zlomů se musí vytvořit v subtelomerických oblastech vkládaného chromozómu a jeden v pericentrické oblasti recipientního chromozómu (recipientní chromozóm v podstatě podstoupí centromerický rozpad) (Schubert & Lysák, 2011).

## 2.6 Polymorfizmy a mechanismy fixace chromozómových přestaveb

Každá strukturální přestavba se nejdříve vyskytne v heterozygotním stavu a v rámci jedné populace se může různě dlouho udržovat jako polymorfizmus. Podle toho, jaký má efekt na životaschopnost a fertilitu svého nositele, se časem buď z populace vyřadí, nebo se rozšíří u většího počtu jedinců v populaci a časem se fixuje v homozygotním stavu (pakliže heterozygotní stav nepřináší nějakou evoluční výhodu, viz dále) (King, 1993). I různé populace stejného druhu se mohou lišit přestavbami, v takovém případě hovoříme o polytypii, chromozómových rasách, cytotypech apod. Učebnicovým příkladem jsou metacentrické rasy u myši domácí, konkrétně poddruh *Mus musculus domesticus* (Garagna et al., 2014). Mnoho polymorfizmů bylo zdokumentováno především napříč savci (Dobigny et al., 2017).

CRs mohou být fixovány buď pozitivní či balancující selekcí (viz kapitola 2.8) nebo neutrálními procesy jako např. náhodným genetickým driftem v malých izolovaných populacích (King, 1993), případně pomocí samičího meiotického tahu souvisejícího s velikostí centromery (Pardo-Manuel de Villena & Sapienza, 2001; Brown & O'Neill, 2010; Clark & Akera, 2021). V tom posledním případě se jedná o tzv. centromerický tah, který ve stručnosti znamená, že pokud u heterozygota pro např. centrickou fúzi má metacentrický chromozóm rozsáhlejší oblast centromery, než je součet centromer nefúzovaných jednoramenných chromozómů, tak [z důvodu velmi komplexních mechanismů, např. (Kumon & Lampson, 2022)] segreguje metacentrický chromozóm preferenčně do vajíčka, zatímco

jednoramenné chromozómy do pólové buňky a s ní pak zaniknou. Tímto způsobem se mohou v karyotypech postupně akumulovat produkty fúzí, nicméně polarita dělicího vřetenka se může časem obrátit (Pardo-Manuel de Villena & Sapienza, 2001; Blackmon et al., 2019). Posledním zajímavým fenoménem je tzv. karyotypová ortoselekcce, tj. jev při kterém skupina blízce příbuzných druhů preferenčně fixuje tentýž druh přestavby, protože jsou k němu náchylní z důvodů podobné komplexity (a případných nedostatků) mechanismů oprav DNA a/nebo podobné distribuce repetitivní DNA (King, 1993).

## 2.7 Strukturní varianty

Termín strukturní varianty neboli SVs, z angl. *structural variants*) je označení do značné míry synonymní pro polymorfizmy chromozómových přestaveb. Dle definice se jedná o variabilitu v úseku genomu zahrnujícím alespoň 50 bp (často mnoho kb až Mb) (Feuk et al., 2006; van Dijk et al., 2018). Jedná se často o jednoduché CRs, ale řadí se sem i komplexní strukturní přestavby, které jsou obvykle nalézány v genomech nádorových buněk (Pellestor et al., 2011). Stejně jako CRs se i SVs dělí na balancované a nebalancované na základě toho, zda se mění množství genetické informace. U nebalancovaných přestaveb (delece a duplikace/amplifikace) se používá označení varianty počtu kopií (neboli CNVs, z angl. *copy number variants*). V tomto případě se jedná o minimálně 1 kb dlouhý úsek DNA, který se liší v počtu kopií genetické informace v porovnání s referenčním genomem stejného druhu (Feuk et al., 2006). Jiným typem genetické variability, který byl donedávna majoritně studován v genomových pracích jsou polymorfizmy záměn (substitucí, inzercí, delecí) na úrovni jednoho nukleotidu, tzv. SNP (z angl. *single-nucleotide polymorphism*) neboli „snipy“. Tento druh variability lze studovat snáze pomocí dlouhodobě optimalizovaných protokolů a algoritmů, zatímco diverzita SVs zůstávala dlouho neprobádána a v pozadí zájmu, dokud nedošlo k rapidnímu rozvoji moderních sekvenačních metod a asociovaných bioinformatických postupů (Sedlazeck et al., 2018a; van Dijk et al., 2018; Mahmoud et al., 2019; Vollger et al., 2019, 2020). Ukázalo se, že SVs jsou podstatně častější a mají větší potenciál směřovat evoluci genomu než „snipy“ (Chiang et al., 2017; Mérot et al., 2020; Yuan et al., 2021). V posledních letech přibývá studií, které ukazují, že strukturní varianty mohou být u různých eukaryot asociovány s komplexními fenotypovými znaky, adaptací na lokální podmínky, reprodukční izolaci a diverzifikaci (viz následující kapitola).

## 2.8 Chromozómové přestavby a jejich úloha v evoluci

Již velmi dlouhou dobu se vedou debaty o tom, jaký význam mají CRs v evoluci druhů. Jednou z možností je, že samotná přestavba vede k fenotypovým změnám, které vylučují páření mezi jedinci lišící se přítomností/absencí dané přestavby (tj. pre-zygotická bariéra), nebo se jedinci spáří, ale vzniká potomstvo, které není životaschopné, nebo má sníženou fertilitu (tj. post-zygotická bariéra) (např. Dickinson et al., 2012; Coughlan & Matute, 2020). Nekompatibilita mezi rodičovskými genomy může

vést k zablokování meiotického dělení v některém z kontrolních bodů, nebo může dojít k chybné segregaci chromozómů nebo jejich částí do gamet, k čemuž může výrazně přispět rekombinace v oblasti přestavby (Harewood & Fraser, 2014; Raudsepp & Chowdhary, 2016; Tuna et al., 2019).

Jaký má u heterozygotů s přestavbou daná přestavba vliv na tvorbu životaschopných gamet souvisí s tím, jestli v oblasti přestavby dojde k rekombinaci a také, jak chromozómy segregují z multivalentů. První zmiňovaný případ se běžně popisuje na příkladu para- a pericentrické inverze. Během prvního meiotického dělení je pro životaschopnost a fertilitu potomků heterozygota s inverzí zásadní, jestli dojde ke crossing-overu (CO) v oblasti invertovaného úseku. Pokud ne (nebo pokud dojde ke dvěma CO), polovina gamet bude přenášet chromozóm s inverzí, ale jejich genetický obsah bude balancovaný. Pokud však dojde k jednomu CO v oblasti inverze, bude vznikat určitá frakce nebalancovaných gamet. V případě pericentrické inverze budou některé gamety obsahovat chromozóm s delecí/duplikací některých segmentů, v případě paracentrické inverze bude navíc docházet ke tvorbě dicentrického a acentrického fragmentu (Lysák & Schubert, 2013). Obě tyto varianty by měly za následek snížení životaschopnosti potomstva, proto je potřeba, aby rekombinaci bylo zabráněno (Navarro et al., 1997).

Hybridní nekompatibilita má zabránit mísení čistých druhů. Nicméně ne vždy musí nutně speciace či diverzifikace souviset primárně s karyotypovými změnami, neb mohou být také způsobeny např. geografickými nebo behaviorálními bariérami (Brown & O'Neill, 2010; Ravinet et al., 2017). Hybridní nekompatibilita je podstatou tzv. tradičních modelů vysvětlujících zásadní vliv CRs v reprodukční izolaci. Tyto tzv. *Hybrid sterility/underdominance models* tedy předpokládají, že u geograficky oddělených populací se nahromadí karyotypové změny, které vedou k redukci fertility a životaschopnosti, pakliže se jedinci z těchto odlišných populací druhotně potkají a spáří se. To má za následek zamezení genového toku mezi jedinci divergujících populací (Brown & O'Neill, 2010). Mezi zásadní nedostatky tradičních modelů patří, že přestavba se silně negativním efektem na reprodukční zdatnost jedince neboli fitness se z populace vyřadí pod vlivem selekce. Může být fixována pouze genetickým driftem v malých populacích. Taková situace ale není v přírodě až tak moc častá. Na druhou stranu přestavba s malým efektem na fitness zase nebude důvodem pro reprodukční izolaci (Navarro & Barton, 2003; Faria & Navarro, 2010).

Novější modely, tzv. „*Suppressed recombination models*“, mají obecnější platnost, protože vysvětlují chromozómovou speciaci i v případech probíhajícího genového toku mezi populacemi. Přestavba nemusí nutně mít silně negativní efekt na fitness. Selektce a/nebo sterické důvody zabraňují rekombinaci v oblasti přestavby u heterozygotů, takže zatímco zbytek genomu rekombinuje, v oblasti přestavby se vytváří tzv. ostrov genetické divergence, který akumuluje genetické změny. Kombinace suprese rekombinace a sekvenční divergence v oblasti přestavby vede k odlišení jedinců mezi různými populacemi (Jackson, 2011). Na rozdíl od tradičních modelů, byly tyto potvrzeny i experimentálně (Faria & Navarro, 2010).

### 2.8.1 Vliv CRs na funkci a evoluci genů

Po představení jednotlivých typů chromozómových přestaveb bych ráda věnovala pozornost jejich konkrétnímu vlivu na funkci a evoluci genů. Chromozómové přestavby mohou v první řadě gen přerušit; ke zlomu většinou dochází v intergenových oblastech nebo intronech. Dále mohou CRs vést ke změně expresního profilu přilehlých genů na základě toho, zda tyto mají nebo již nemají ve své blízkosti svoji původní regulační oblast (Harewood & Fraser, 2014). Daný gen se může vlivem přestavby dostat do oblasti vlivu jiných regulačních sekvencí a/nebo se může dostat do oblasti s odlišným epigenetickým profilem (např. jeho exprese bude umlčena přilehlým heterochromatinem) (Goettel & Messing, 2010; Muñoz & Sankoff, 2012; Harewood & Fraser, 2014). Na evoluci nových genů má dále zásadní vliv genová duplikace, amplifikace, a velmi často v evoluci i polyploidizace, neboť duplikáty genů nejsou pod silnou selekcí jako jejich originály, a tedy mohou buď degenerovat anebo dát vzniknout nějaké důležité evoluční novince (recentní přehled viz Heslop-Harrison et al., 2023).

Přestavba může mít také zásadní vliv na změnu pozice chromozómu v jádře. U většiny eukaryot se ve středu jádra nacházejí hojně transkribované chromozómy nebo jejich segmenty/geny, kdežto u jaderné membrány, kde se formuje heterochromatin, se soustřeďují umlčené, transkripčně inaktivní oblasti genomu (Towbin et al., 2013). Každý chromozóm v interfázním jádře zaujímá konkrétní oblast (teritorium), a její poloha se výrazně mění v průběhu vývoje a je rovněž tkáňově specifická. To znamená, že v různých tkáních spolu sousedí různé chromozómy, mezi kterými může docházet k přestavbám (shrnuto např. v Cremer & Cremer, 2010; Misteli, 2020). Takové přestavby mohou pak souviset s tkáňově specifickou nádorovou transformací, která může být např. spjata s tvorbou fúzních neboli chimérických genů (tj. např. po chromozómové translokaci se spojí fragmenty dvou genů a začne se exprimovat protein s novými vlastnostmi) (Tuna et al., 2019). Translokace nebo fúze fyzicky spojuje do jedné vazebné skupiny genetický materiál dříve přítomný na různých chromozómech a tím také mění profily rekombinační četnosti napříč genomem (Vara et al., 2021; Yoshida et al., 2023). Výrazný efekt na evoluci genů může mít i centrický rozpad (např. (Fan et al., 2019). Přestavby mohou ale nemusí mít výrazný vliv na změnu trojrozměrné struktury genomu, na profily kontaktních frekvencí mezi chromozómy a na epigenetický (aktivní vs. Inaktivní) stav chromatinu (např. Vara et al., 2021; Yin et al., 2021; X. Li et al., 2023). V oblasti přestavby, a dokonce i poměrně daleko od zlomového místa, bývá potlačena nebo výrazně redukována rekombinace, u fúzí a translokací (na rozdíl od inverzí) dokonce nejen u heterozygotů (Wellband et al., 2019), ale i u homozygotů pro přestavbu (D. Liu et al., 2018; Vara et al., 2021; Yoshida et al., 2023). To znamená, že alely některých sousedících genů se ocitnou v těsné vazbě a v této sestavě segregují spolu opakovaně do dalších generací (je mezi nimi tzv. vazebná nerovnováha). Pokud tyto konkrétní alely sousedících genů přinášejí nějakou selektivní výhodu/vyšší fitness pro své nositele, tak jsou v populaci udržovány jako tzv. „supergen“ a obvykle se udržují v polymorfním stavu v dané populaci díky tzv. balancující selekci (přehled viz Gutiérrez-Valencia et al., 2021) Supergeny mohou např. podpořit lepší adaptaci dané populace na konkrétní podmínky

prostředí (vytvářejí se různé ekotypy), nebo mohou podpořit vývoj komplexních fenotypových znaků (např. mimikry) či evolučních strategií. Různé populace stejného živočišného nebo rostlinného druhu tak mohou dojít do bodu reprodukční izolace i přes to, že by mezi nimi dosud mohl probíhat genový tok (Wellenreuther & Bernatchez, 2018; Wellband et al., 2019; Z. Liu et al., 2022).

## 2.9 EBRs a fragilní místa

Základním předpokladem vzniku většiny CRs je vznik zlomu v DNA. I přes původní hypotézy o vzniku náhodných zlomů v genomu, bez preference pro konkrétní oblasti (tzv. „*random breakage model*“; Nadeau & Taylor, 1984), se na základě četných pozorování ukázalo, že v určitých oblastech genomu dochází ke dvouvláknovým zlomům opakovaně (Pevzner & Tesler, 2003; Alexeev & Alekseyev, 2017). Tyto oblasti dostaly označení evolučně konzervované dvouvláknové zlomy (neboli EBRs, z angl. *evolutionary breakpoint regions*) a mohou představovat i shluk několika přilehlých alternativních míst náchylných ke zlomu (Sankoff, 2009; Berthelot et al., 2015). Zůstává otázkou, jestli jejich přítomnost nesouvisí spíše s tím, že zlomy v jiných oblastech jsou vysoce škodlivé z hlediska fitness a vyřadí se z populace selekcí (Larkin et al., 2009). Konzervativní oblasti, kde obvykle nedochází k přerušení sekvence DNA, se nazývají HSBs (homologní syntenní bloky, z angl. *homologous syteny blocks*). Syntenie znamená, že se dané sekvence (typicky lokusy dvou či více genů) nacházejí spolu na stejném vlákně DNA (ve stejné vazebné skupině) a konzervovaná syntenie znamená, že je toto uspořádání zachováno v genomech příbuzných druhů (Ehrlich et al., 1997; Damas et al., 2021). Na úrovni analýzy somatických tkání jedince se pro oblasti opakovaných zlomů používá termín „fragilní místa“, která obvykle vedou k patologickým přestavbám (Wilhelm et al., 2018). V genomu se EBRs vyskytují především v místech bohatých na GC páry, s vysokou hustotou genů, replikačních počátků, repetitivních sekvencí a segmentálních duplikací, v hypometylovaných oblastech a v hraničních oblastech tzv. topologicky asociovaných domén (TADs, z angl. *topologically associating domains*) (Berthelot et al., 2015; Damas et al., 2021; Álvarez-González et al., 2022), což jsou smyčkové domény, které organizují chromatin chromozómů v interfázním jádře (přehled viz Misteli, 2020). Tato fakta naznačují, že samotná struktura genomu hraje podstatnou roli, resp. předurčuje výskyt chromozómových přestaveb.

## 3 Metody zkoumání CRs

Strukturní přestavby či zde dosud nejmenované numerické chromozómové změny (tj. ztráta nebo nabytí jednotlivých chromozómů neboli aneuploidie) lze primárně analyzovat konvenční a molekulární cytogenetikou. Tyto metody jsou koneckonců vhodné i pro počáteční ověření počtu vazebných skupin a vyloučení možných chromozómových anomálií u jedinců určených pro sekvenování genomu (Deakin et al., 2019). Při detailnější analýze celkového spektra přestaveb v rámci srovnávaného souboru příbuzných druhů jsou ovšem molekulárně cytogenetické metody technicky i časově velmi

náročné a nedokážou zachytit úplně všechny strukturní změny, především ne změny malého rozsahu, např. drobné delece nebo paracentrické inverze (Damas et al., 2021; Rhie et al., 2021; Dutta et al., 2022). Detekce chromozómových přestaveb je dále možná na bioinformatické úrovni. K tomu, aby bylo možno precizně identifikovat a charakterizovat oblasti, kde došlo ke dvouvláknovým zlomům vedoucím k přestavbám, je potřeba mít k dispozici kvalitně poskládané genomy a vhodně vybrané bioinformatické algoritmy (Kosugi et al., 2019), a to není úplně jednoduché. Např. součástí projektu 1000 Genomes Project byla detekce SVs čtyřmi metodologickými přístupy, přičemž všechny poskytly v tomto ohledu odlišné výsledky (Durbin, Altshuler et. al, 2010; Mills et al., 2011). Detekce EBRs a SVs patří mezi nejnáročnější úkoly v oblasti genomiky. Konkrétně SVs se často vyskytují v heterozygotním stavu a na jejich okrajích se mohou vyskytovat vysoce identické repetitivní sekvence, které znesnadňují správné poskládání této oblasti (Treangen & Salzberg, 2012). Míra obtížnosti nalezení EBRs a SVs dále souvisí s rozsahem přestavby a také s velikostí a obsahem repetitivní DNA studovaného genomu (Peona et al., 2018, 2021).

### 3.1 Cytogenetika a komparativní cytogenomika

Konvenční cytogenetika, tj. barvení chromozómů jadernými barvivy (např. Giemsa) a pruhovací techniky, stále patří mezi rutinní postupy pro odhalování strukturních přestaveb, obzvláště pak u savců a člověka, kde konkrétně pruhovací techniky poskytují konzistentní vzory pruhů umožňující identifikaci homologních chromozómů a také celou řadu strukturních změn/anomálií (Trask, 2002). Nicméně u celé řady skupin organismů většina pruhovacích technik nefunguje tak jako u teplokrevných obratlovců, protože jejich genomy nejsou organizovány do AT- a GC-bohatých segmentů (shrnuto v Rábová et al. 2015). Proto je potřeba především u nich použít pokročilejší cytogenetické techniky (viz dále). I přes rapidní rozvoj sekvenačních technologií, které dokáží odhalit přestavby menšího rozsahu, však karyotypování zůstává nedílnou součástí procesu ověřování výsledků genomických metod (Rhie et al., 2021).

Možnosti použití cytogenetických technik se výrazně rozšířily s nástupem metody fluorescenční hybridizace *in situ* (FISH, z angl. *fluorescence in situ hybridization*) (Pardue & Gall, 1969), která umožňuje fyzické mapování vybrané sekvence přímo na chromozómech za použití hybridizační sondy (tj. krátký úsek, obvykle do 500 bp, případně směs takových fragmentů pokrývajících delší úseky chromozómu), která je následně detekovatelná přímo na chromozómech jako fluorescenční signál. Tímto způsobem lze s různou mírou preciznosti (v závislosti na variantě FISH) detekovat homologní úseky chromozómů a jejich přestavby (přehledy viz, Levsky & Singer, 2003; Huber et al., 2018; Hu et al., 2020).

Pro analýzu míry zachování konzervované syntenie mezi více či méně příbuznými druhy a tím pádem i pro mapování EBRs a SVs se v rámci komparativní cytogenomiky využívají především metody WCP (z angl. *whole chromosome painting*) a BAC-FISH (BAC je zkratka pro *bacterial artificial*

*chromosome*). Při WCP se používají celochromozómové malovací sondy připravené z izolovaných chromozómů získaných pomocí sortování při průtokové cytometrii (Doležel et al., 2014) nebo pomocí laserové či jehlové mikrodisekce (Iannucci et al., 2021). V případě BAC-FISH se jedná o fragmenty genomu (o délce 100-500 kbp) zaklonované v rámci přípravy genomové knihovny (v období před nástupem sekvenování druhé generace) do vysokokapacitních vektorů zvaných bakteriální umělé chromozómy (Shizuya & Kouros-Mehr, 2001). Pro mezidruhovou aplikaci metod WCP a BAC-FISH (tj. sonda z jednoho druhu je hybridizována na chromozómy jiného druhu) se souhrnně používá označení Zoo-FISH (např. Raudsepp et al., 1996; Ferguson-Smith & Trifonov, 2007). Díky tomuto postupu můžeme analyzovat pozici homologních úseků a přeskupování syntenních bloků v průběhu evoluce, pakliže známe fylogenetické vztahy dané taxonomické skupiny (Raudsepp & Chowdhary, 2001; Ferguson-Smith & Trifonov, 2007; Damas et al., 2017). Nevýhodou této metody je pak její pracnost, složité porovnávání hybridizačních vzorů napříč druhy a možné potíže při porovnávání evolučně vzdálených druhů (tj. při výraznějším přeuspořádání syntenních bloků již hybridizační sondy na vzdálené druhy nemusí fungovat) (Ferguson-Smith & Trifonov, 2007), což moderním sekvenačním metodám nečiní potíže (Hu et al., 2020). Nicméně, na druhou stranu, izolace celých chromozómů umožňuje redukovat komplexitu sekvenování [tj. každý chromozóm se sekvenuje separovaně a my tak víme, že získané sekvence patří k této vazebné skupině; (Iannucci et al., 2021)], což je obzvláště přínosné při sekvenování organismů s velkými (často polyploidními) genomy (Doležel et al., 2014). Navíc, uvedené metody mohou sloužit jako kontrola správného poskládání genomu. Jako příklad lze uvést BAC-FISH úseků genomu prasete (O'Connor et al., 2017) a dvou druhů želv (Lee et al., 2020), která v obou případech odhalila chybné poskládání určitých úseků genomu.

V současné době je pro metody WCP a BAC-FISH slibnou variantou tzv. Oligo-paint FISH, při které se navrhuje sada oligonukleotidových hybridizačních sond *in silico* na základě genomových dat. Tento přístup vytváří vysoce specifické sondy, neboť je možné (a přímo žádoucí) při navrhování sond vynechat repetitivní oblasti, které by jinak mohly vytvářet nadbytečné signály a zkreslovat tak interpretaci výsledků (Jiang, 2019; G. Liu & Zhang, 2021).

## 3.2 Genomické metody

### 3.2.1 Sekvenování druhé generace

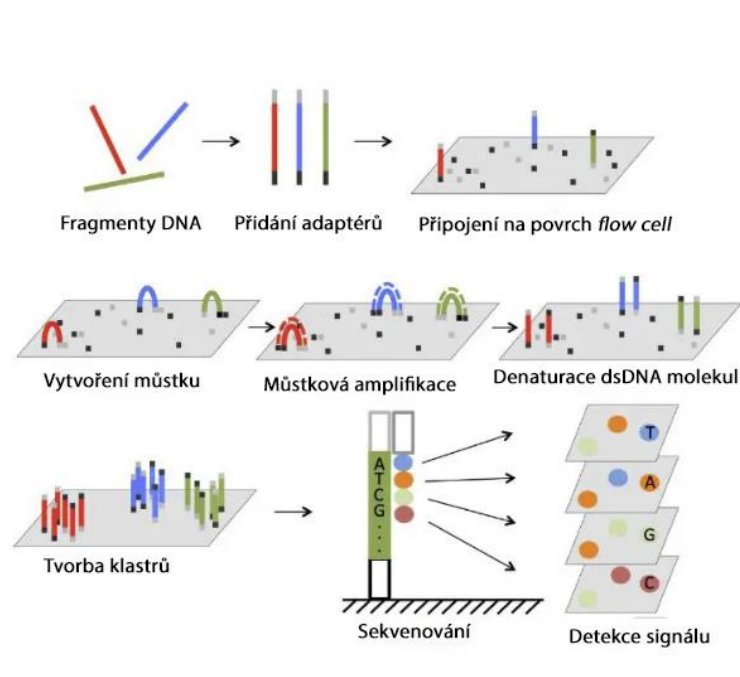
Sekvenování, tj. určování pořadí bází ve zkoumané molekule DNA, začalo být prováděno pomocí dvou pionýrských metod ke konci sedmdesátých let rincipi století (Sanger & Coulson, 1975; Maxam & Gilbert, 1977). Z nich především tzv. Sangerova metoda byla dále metodicky vylepšována, aby bylo jednak možné přečíst co nejdelší úsek DNA na jeden „záťah“, a jednak aby bylo možné proces automatizovat pomocí kapilární elektroforézy v tzv. sekvenátorech (Schloss, 2008). I přesto ale tato metoda zůstala časově i finančně náročná, a tak po dvou dekadách přišly na trh tzv. metody sekvenování druhé generace (neboli NGS, z angl. *Next generation sequencing*) (přehled viz Metzker, 2010).

NGS je souhrnné označení pro automatizované metody, které zahrnují masivní paralelní sekvenování celého souboru (knihovny) malých fragmentů (majících uniformní délku, často kolem 450 bp) zahrnujících dohromady celý zkoumaný genom. Společným principem těchto metod je nutnost amplifikace templátových molekul DNA pomocí polymerázové řetězcové reakce (neboli PCR, z angl. *polymerase chain reaction*). Fragментy DNA jsou ukotveny prostřednictvím adaptorů na pevném povrchu mikrodestičky (čipu), případně u některých metod probíhala amplifikace templátu jiným způsobem (např. v emulzi vody a oleje u pyrosekvenování). Důvodem amplifikace templátu je, aby bylo možné při následném sekvenování přístrojem zaznamenat signál při připojení nukleotidu s fluorescenční značkou (pouze shluk nakopírovaných fragmentů vydá detekovatelný signál). Fragментy jsou následně čteny v procesu opakujících se cyklů sekvenování a odmyváání a je obvykle chemicky zajištěno, aby se v jednom cyklu připojil pouze jeden komplementární fluorescenčně značený nukleotid, zatímco přebytečné nekomplementární nukleotidy jsou odmyty.

Mezi metody druhé generace patří pyrosekvenování (Roche) a principiálně podobný Ion Torrent, sekvenování ligací (SOLiD) a sekvenování pomocí syntézy (Solexa/Illumina). Postupně se na trhu osamostatnila poslední jmenovaná metoda, která po svém uvedení roku 2004 firmou Solexa byla o dva roky později převzata společností Illumina a odtud pochází její současné označení.

Celý proces metody Illumina (popsaný též na Obr. 5) se skládá ze čtyř hlavních kroků: i) příprava knihovny stejně dlouhých fragmentů DNA pomocí automatického sonikátoru a navázání adaptorů majících sekvenci komplementární k oligonukleotidům, jimiž je poset celý povrch sekvenačního čipu (*flow cell*), ii) vytváření klastrů templátových fragmentů pomocí tzv. můstkové amplifikace (tj. výše zmíněná amplifikace PCR na pevném podkladu), iii) sekvenování principem cyklické reverzibilní terminace (viz dále) a iv) analýza dat. Princip sekvenování (tj. cyklické reverzibilní terminace) spočívá v navázání deoxyribonukleosidtrifosfátů (dNTPs), které nesou fluorescenční značku a také blokující chemickou skupinu na 3'OH skupině deoxyribózy. Když dojde k zařazení nukleotidu polymerázou, zastaví se syntéza a může se změřit fluorescence. Následně se přidá činidlo, které odstraní fluorescenční i blokační skupinu (tj. reverzibilní terminátor) a syntéza může pokračovat přidáním dalšího značeného dNTP (Metzker, 2010).

Největší přínos sekvenování Illumina spočívá ve schopnosti paralelního sekvenování obrovského množství fragmentů za současné velmi nízké chybovosti (dosahující až 99 %). Oproti Sangerovu sekvenování se také rapidně snížily náklady na přípravu i samotné sekvenování (Mardis, 2008).



**Obrázek 5:** Postup jednotlivých kroků metody Illumina, upraveno podle internetového zdroje č. 1

Sekvenování metodou Illumina má na druhou stranu největší limity v délce vygenerovaných sekvencí. Délka sekvenčního čtení neboli „readu“ je často pouhých 100–150 bp, což limituje použití této metody pro sestavování genomů *de novo*. Pokud je ale k dispozici jiný již poskládaný genom téhož či blízce příbuzného druhu, je možné tento genom použít jako tzv. referenci a krátká sekvenční čtení na něj mapovat (tzv. reference-guided assembly) (Mérot et al., 2020). Jedním z hlavních důvodů limitace krátkých sekvenčních čtení je fakt, že není možné správně poskládat dlouhé úseky repetitivních sekvencí. Tento nedostatek částečně kompenzuje fakt, že fragmenty DNA jsou obvykle čteny z obou konců (tzv. sekvenování párových konců, z angl. *paired-end sequencing*), což usnadňuje přemostění repetitivních úseků, neb prostřední nesekvencovaný úsek je vždy stejně dlouhý, a tak můžeme správně seřadit párové „ready“ do kontinuálních úseků (kontigů). I přes tuto strategii ovšem není možné poskládat vysoce repetitivní části genomů, neb nemůžeme říct jistě, k jaké vazebné skupině tyto úseky patří a v jaké jsou orientaci, a tak jsou odpovídající „ready“ obvykle z procesu poskládání genomu (z angl. *genome assembly*) vynechány (Metzker, 2010; Schadt et al., 2010). Illumina tak nedokáže detekovat chromozómové přestavby, pouze bodové mutace (SNPs, indely) (English et al., 2015). Není proto dostačující pro komplexní analýzu syntenních bloků i EBRs (Damas et al., 2021). Výjimkou je sekvenování knihovny fragmentů, tzv. *mate-pair sequencing*, které s větší pravděpodobností překlene oblasti SVs a EBRs (Lucek et al., 2019).

Dalším nedostatkem metody Illumina je amplifikační bias, které má vliv na zkreslení výsledků (Whiteford et al., 2009; Alkan et al., 2011). Pro kontrolu validity dat se postupně vyvíjely další systémy, např. „spike-in“. Jedná se o přidané fragmenty se známou sekvencí, které jsou analyzovány společně se sekvenovanými daty, čímž se mohou odhalit chyby nebo zkreslení výsledků (Dunning et al., 2008).

### 3.2.2 Sekvenování třetí generace

Zásadní změnou u metod sekvenování třetí generace je použití jednotlivých dlouhých nativních molekul DNA jako templátu. Není zde zahrnut proces amplifikace fragmentů pro účely detekce signálů, a tudíž nedochází ani k amplifikačnímu biasu. Další obecnou a poměrně zásadní výhodou vyplývající z vlastnosti templátové DNA je délka sekvenovaných úseků, která je mnohonásobně vyšší než u metod předchozí generace, čímž se může zvýšit pravděpodobnost detekce neobvyklých variant v genomu pozorovaných např. v oblastech akumulace repetitivních sekvencí (Schadt et al., 2011; Srinivasan & Batra, 2014; Gao et al., 2023; Liao et al., 2023). Největším úskalím metod třetí generace je poměrně vysoká chybovost (viz dále).

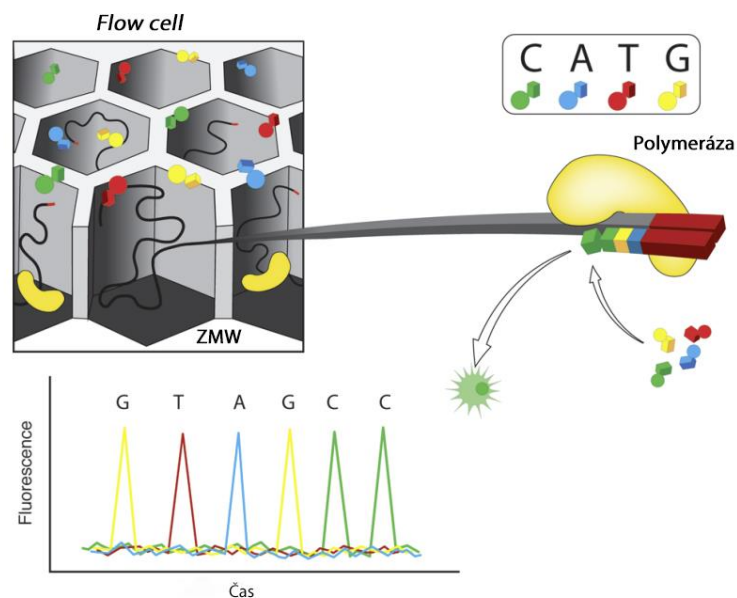
V následujících kapitolách představím současné technologie sekvenování třetí generace, zhodnotím jejich výhody a nevýhody při zkoumání strukturních variant, chromozómových přestaveb a EBRs.

#### 3.2.2.1 PacBio – SMRT

SMRT (sekvenování jedné molekuly v reálném čase, z angl. *single molecule realtime sequencing*) je metoda, kterou vyvinula společnost PacBio (Pacific Biosciences). Princip metody spočívá ve sledování jedné molekuly polymerázy přímo během procesu syntézy vlákna DNA, kdy přichází fluorescenčně značené nukleotidy tento proces nezastavují (Eid et al., 2009). Prvním krokem je příprava tzv. SMRT bell knihovny, což se udělá tak, že se na molekulu DNA navážou adaptérové molekuly. Následně na adaptéry nasedne DNA polymeráza s primerem a syntéza se pozoruje v reálném čase.

Pozorování jedné molekuly s sebou ale nese překážku, jak detekovat fluorescenci pouze jednoho nukleotidu v reálném čase za přítomnosti velkého počtu nukleotidů v celkové směsi, které mohou vytvářet výrazné fluorescenční pozadí. Tento problém je možné odstranit snížením sledovaného objemu, a proto byly vyvinuty čipy s jamkami o velice malém průměru (ZMW; z angl. *zero-mode waveguides*) (Levene et al., 2003; McCarthy, 2010). Jedná se o póry o průměru 10 nm vyrobené z kovové vrstvy o tloušťce 100 nm nanesené na skleněném substrátu. V čipu SMRT cell se nacházejí desítky tisíc ZMWs. Světelným signálem jsou osvětlovány ZMWs zespodu takovou vlnovou délkou (kolem 100 nm), aby nebylo umožněno prostoupení světla na povrch SMRT cell. Při průchodu světla skrze ZMW nedochází přímo k průchodu světleného signálu, jako spíše k jeho exponenciálnímu rozpadu, proto je osvětleno pouze spodních 30 nm jamky (Schadt et al., 2010).

Samotná detekce přidávaných nukleotidů při syntéze vypadá následovně: V každé jamce se nachází právě jedna DNA pol, která je přichycena ke spodnímu povrchu skleněného substrátu (tj. do centra osvětlení). Přichycení je zprostředkováno vazbou biotinu (na DNA polymeráze) na streptavidin (spodní sklíčko ZMW) (Eid et al., 2009). DNA polymeráza zachytí jednovláknovou templátovou molekulu DNA a začne syntéza. Nukleotidy se vznášejí nad prostorem jamky v požadované koncentraci a na rozdíl od metod druhé generace jsou fluorescenčně značeny na svém terminálním (gamma) fosfátu, a ne na bázi. Při syntéze putují fluorescenčně značené nukleotidy dolů k polymeráze (na úroveň 30 nm) a během správného komplementárního párování vydrží v komplexu s DNA polymerázou dostatečně dlouho na záznam fluorescence. Tím, že se fluorescenční značka nachází na gama fosfátu, dojde po začlenění nukleotidu k jejímu odštěpení a detekovatelný signál se tím přeruší (Eid et al., 2009) (Obr. 6).



**Obrázek 6:** Znázornění procesu sekvenování probíhajícího na flow cell, upraveno podle Escalante et al. (2014).

Metoda SMRT je v současnosti jednou z nejpoužívanějších sekvenačních metod. Její hlavní výhodou je omezení časové náročnosti, ať už přípravy vzorku nebo i samotného sekvenování. Výsledek je možné dodat již v řádu minut v porovnání s metodou Illumina, při jejímž použití trvá analýza i několik dní. Díky možnosti paralelního pozorování DNA polymerázy v ZMWs je umožněno čtení delších sekvencí. Výsledné sekvence dosahují délky 10 000 bp, maximální délka čtení je 100 kb. Metoda umožňuje sestavení genomu *de novo*, a používá se i pro poskládání mateřské haploidní sady chromozómů, která nemá promíchané ready s otcovskou haploidní sadou (tzv. fásování a „*haplotype-resolved genomes*“), viz kapitola 3.4 (Schadt et al., 2011; Gao et al., 2023; Liao et al., 2023).

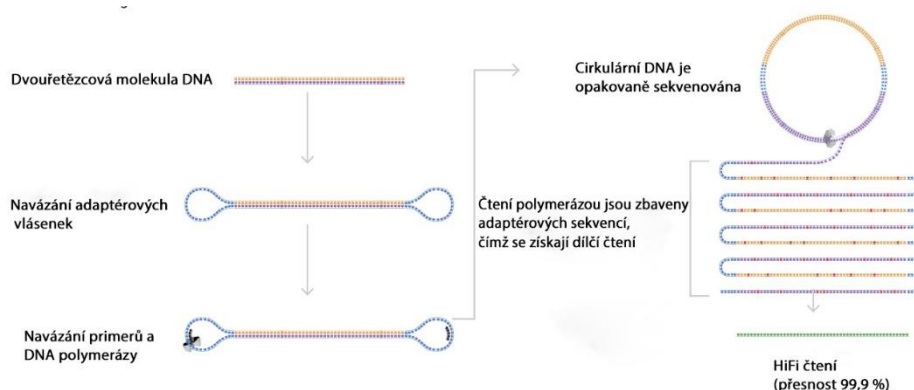
I přes to, že metoda PacBio SMRT má spoustu výhod, je třeba zmínit i některé její nedostatky. Chybovost sekvenování může dosáhnout až 5 % (Schadt et al., 2010; Carneiro et al., 2012). K chybám s určováním delecí a insercí může docházet při chybném zachycení signálu během sekvenace, např. když se dva nukleotidy začlení příliš rychle po sobě, může se do výsledků projevit jako delece (Eid

et al., 2009). V porovnání s metodami druhé generace má SMRT i nižší výkon, protože je omezená počtem ZMWs, které lze číst najednou (Rhoads & Au, 2015).

V kontextu chromozómových přestaveb bylo PacBio také velkým průkopníkem, a to především z důvodu produkce dlouhých čtení oproti metodám NGS. Pro zpracování dat, ve kterých se mohou potencionálně nacházet různé strukturní varianty, bylo vyvinuto detekční zařízení MultiBreak-SV, který vytváří mnoho alternativ zarovnání pro dlouhé čtecí sekvence, čímž nabízí velký čtecí rámec pro detekci SVs (Ritz et al., 2014). Vzhledem ke své chybovosti se nepoužívá tato metoda samostatně příliš často a je zapotřebí kombinace s jinými technologiemi, které produkují přesnější data (Chaisson et al., 2015). Pro detekci chromozómových přestaveb, konkrétně oblastí zlomů a znovuspojení na chromozómech, se s úspěchem využívá např. HiFi přístup u metody PacBio.

### 3.2.2.2 HiFi přístup u PacBio

Vzhledem k nepřesnosti, kterou produkují CLR (souvislé dlouhé čtení, z angl. *continuous long reads*) přístupy u SMRT technologie (i nanopórového sekvenování), bylo potřeba vyvinout přístup, který bude mít výrazně nižší chybovost. Společnost PacBio pro tyto účely vyvinula technologii CCS (kruhové konsenzuální sekvenování, z angl. *circular consensus sequencing*) neboli HiFi (vysoká přesnost, z angl. *high fidelity*). Tento přístup generuje dlouhé sekvence s přesností až 99,5 %, přičemž délka sekvencí je vyšší než 10 kb. Jeho hlavní výhodou je sekvenování cirkulární molekuly, které se



**Obrázek 7:** Postup HiFi čtení, upraveno podle internetového zdroje č. 2

několikrát po sobě opakuje a srovnání výsledných osekvenovaných úseků umožňuje identifikovat a vyřadit chybně přečtené úseky. Cirkularizace DNA probíhá prostřednictvím navázaných adaptérových molekul (Obr. 7).

Výhodou přístupu HiFi jsou snížené časové i finanční náklady při zarovnání čtení a sledování překryvů, dále je možné pro tuto metodu používat stejné programy jako při analýze krátkých sekvencí, např. GATK (z angl. Genome Analysis Toolkit), který byl navržen pro detekci strukturních variant. Přístup HiFi zkrátka spojuje to nejlepší ze sekvenování druhé a třetí generace (Wenger et al., 2019).

Technologie HiFi může sama o sobě odhalit různé CRs, aniž by byla doplněna o další technologie jako jsou optické nebo kontaktní mapy chromozómů (Rabanal et al., 2022). Může být použita i při *de novo* sestavování genomů (Li et al., 2021).

### 3.2.2.3 Nanopórové sekvenování

Druhou nejčastěji používanou sekvenační metodou generující dlouhé sekvence je nanopórové sekvenování, které se od předchozích metod liší tím, že přímo detekuje nukleotidy bez aktivní syntézy DNA. Princip technologie byl navržen již roku 1996, kdy byl popsán systém průchodu jednořetězcové molekuly DNA (ssDNA, z angl. *single-strand DNA*) skrze biologický nanopór. Autoři studie předpokládali, že jednotlivé nukleotidy vyvolávají při průchodu nanopórem charakteristické změny toku iontů. V konečném důsledku tedy zachycení změn iontového prostředí vede k určování sekvencí DNA (Kasianowicz et al., 1996).

Kritickým krokem nanopórového sekvenování je příprava genomové knihovny. Tento proces vyžaduje extrakci neporušené vysokomolekulární DNA o délce 10-30 kb (Jain et al., 2018). Proteinový nanopór je stabilizován v elektricky odolné polymerní membráně a prochází skrze něj vysokomolekulární jednovláknová DNA. Pomocí nastavení napětí na této membráně senzory detekují změny toku iontů způsobené obsazením nanopóru jednotlivými nukleotidy v reálném čase, při průchodu molekuly DNA. Sekvenační komora naplněná elektrolytem je rozdělena na dvě části: *cis* a *trans*. Tyto části jsou odděleny biologickou membránou, ve které se nachází nanopóry. Součástí aparatury je i elektroda, skrze kterou se do komory zavádí proud a detekuje se změna iontového proudu (Branton et al., 2008).

První sekvenační zařízení pojmenované MinION vyvinula společnost ONT (*Oxford Nanopore Technologies*). MinION se od té doby dále zdokonaloval a v současnosti využívá biologické nanopóry z *Escherichia Coli*, Curlin sigma S-dependent grow (CsgG). Zařízení MinION nabízí možnost sekvenace ssDNA vedoucího i komplementárního vlákna (bez nutnosti ligace), díky čemuž dosahuje až o 5 % vyšší přesnosti než jeho předchozí verze schopné číst pouze jedno vlákno, protože srovnání obou vláken může odhalit případné chyby. Takové sekvenování se označuje jako 1D<sup>2</sup> (Rang et al., 2018). Od roku 2016 se hojně používá od té samé firmy vyvinuté sekvenační zařízení PromethION, jehož aparatura byla navržena jako stolní zařízení. Je tedy výrazně větší, což umožňuje větší počet nanopórů, díky čemuž se může paralelně sekvenovat více molekul DNA. V konečném důsledku PromethION produkuje data s vyšší přesností, neboť lépe odhaluje chyby. Umožňuje mu to jednak vyšší počet, ale také vyšší kvalita pórů, zároveň generuje delší sekvence, u kterých se při porovnávání s referenčním genomem odhalí chyby snáz (De Coster et al., 2019; Kono & Arakawa, 2019).

Hlavní výhodou nanopórového sekvenování je jeho nízká cena oproti v současné době používaným technologiím od společnosti PacBio. Díky tomu, že metoda není závislá na syntéze DNA, je téměř neomezená délka generovaných sekvencí (Jain et al., 2018). Díky ONT a možnosti ULR (z angl. *ultra long reads*) byly překonány mnohé překážky, jako např. dosud neposkládané oblasti vysoce

repetitivních sekvencí lidského genomu, což vedlo k publikování kompletní sekvence lidského genomu a chromozómu X (Miga et al., 2020; Nurk et al., 2022). Dokonce byl osekvenován i genom bahníka australského (*Neoceratodus forsteri*), který má 14x větší velikost než genom lidský a 90% jeho sekvencí tvoří repetice (Meyer et al. 2021). Metoda ULR také vyžaduje pouze minimální zapojení biologických složek (ligázy, DNA polymerázy, dNTPs), čímž se snižuje cena i časová náročnost přípravy.

Stejně jako u metody SMRT (PacBio) je i u ONT největší nevýhodou vysoká chybovost ve srovnání se metodami NGS. V konkrétních číslech se chybovost nanopórového sekvenování pohybuje kolem 1 % v závislosti na tom, o jaké přídatné programy byla původní technologie obohacena (Ni et al., 2023). Na rozdíl od SMRT se však tyto chyby jeví jako systematické a lze je opravit dodatečnou sekvenací krátkých úseků pro pozdější porovnání. Ve studii sekvenování lidského genomu nanopóry roku 2018 byla data opravována pomocí Illumina HiSeq a přesnost dosahovala až 99,96 % (Jain et al., 2018).

Největší přínos ONT při detekci SVs je bezpochyby ve schopnosti generovat ULR, díky čemuž je možné identifikovat přestavby velkého rozsahu. Potenciál odhalování SVs byl testován a výsledky ukázaly, že na podkladu dat získaných nanopórovým sekvenováním je možné identifikovat rozsáhlé delece, inverze i translokace (Norris et al., 2016).

#### 3.2.2.4 10x Genomics Chromium

Tato technologie se vyznačuje analýzou jednotlivých buněk s vysokou přesností a citlivostí. Základním principem je rozdělení jednotlivých buněk na tisíce nebo dokonce miliony mikrokapek, které jsou specificky označeny a jejich obsah je následně sekvenován. Technologie 10x Genomics generuje sekvence o délce 100 kb. Nejčastěji se využívá k detekci krátkých strukturních variant a také napomáhá fázování genomu (Sedlazeck et al., 2018b).

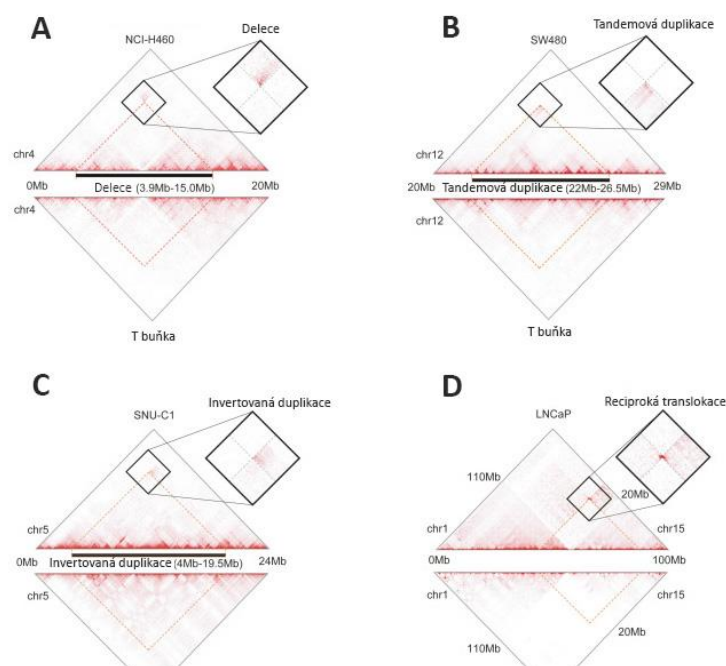
### 3.2.3 Metoda Hi-C

Metoda Hi-C (*High-throughput chromosome conformation capture*) byla vyvinuta přímo pro mapování prostorového uspořádání chromatinu v interfázním jádře. Umožňuje identifikovat fyzické interakce mezi různými oblastmi genomu, což nejen poskytuje informace o kontaktních frekvencích mezi jednotlivými chromozómy, ale také pomáhá stanovit počet vazebných skupin v genomu (Pal et al., 2019). Jedná se tedy o vhodnou metodu interaktomiky a zároveň o doprovodnou metodu pro precizní sekvenování genomu na úrovních „chromosome-level assembly“ a „telomere-to-telomere (T2T) assembly“, a tedy se uplatnila (stejně jako např. optické mapování, viz dále) při kompletním osekvenování lidského genomu (Nurk et al., 2022). Krom jiného tato metoda podpořila platnost modelu chromozómových teritorií a umožnila definovat TADs jako organizační úroveň chromozómů v jádře (Lieberman-Aiden et al., 2009).

Obecný protokol metody Hi-C začíná zesíťováním chromatinu *in situ* např. formaldehydem, což zajistí zafixování trojrozměrné struktury, a tedy i nativních interakcí chromozómů v daném momentu a

dané tkáni. Následně jsou chromozómy štěpeny na kratší fragmenty, dojde k naředění roztoku a koncové sekvence fragmentů držených nyní pouze formaldehydem jsou následně preferenčně ligovány pomocí DNA ligázy. Ke štěpení dochází za pomoci restriktivních endonukleáz, přičemž některé restriktázy nejsou dokonale specifické a mají tzv. hvězdičkovitou aktivitu, tj. jsou schopny štěpit cílové DNA sekvence nejen na svém hlavním rozpoznávacím místě, ale také na dalších místech s podobnou nebo blízkou sekvencí. Řešením tohoto problému je použití kombinace více restriktivních enzymů, nebo použití enzymů, které štěpí DNA více rovnoměrně jako je Dnáz I (Gridina et al., 2021) nebo mikrokokální endonukleáza (Goel & Hansen, 2020). Před následným procesem ligace jsou všechny genomové fragmenty označeny biotinem, díky čemuž se označí ligační spoje. Po ligaci DNA ligázou se v dalším kroku z DNA odstraní formaldehyd nebo jiné použité zesilovací činidlo. Vznikne nám směs mnoha fragmentů DNA, které představují zaligované kousky genomu, které byly v přímém kontaktu. Díky označení biotinem a selektivní vazbou na streptavidin je celá knihovna fragmentů tvořena výhradně produkty ligace. Tyto fragmenty se následně sekvenují platformou Illumina při vysokém pokrytí genomu (např. 50x). (Belton et al., 2012; Wingett et al., 2015; J. Li et al., 2023).

Hlavní nevýhodou metody Hi-C dosud bylo, že neposkytuje přesné informace o uspořádání chromatinu na úrovni jednotlivých buněk nebo malých buněčných populací, protože se jedná se o průměr kontaktních frekvencí všech buněk ve vzorku. Nicméně tzv. single-cell modifikace již jsou dostupné (Chi et al. 2022). Metoda Hi-C dále nemusí být přesná při analýze velkých genomů s vysokým podílem repetitivní DNA. V takovém případě metoda Hi-C neposkytuje plné pokrytí genomu. Další nevýhodou je závislost



**Obrázek 8:** Na vytvořených optických mapách byla sledována frekvence chromozómových kontaktů. Ve spodním trojúhelníku jsou zobrazeny kontroly (T buňka a IMR90), upraveno podle K. Kim et al. (2022).

metody na použití látek jako je formaldehyd, které mohou způsobit zkreslení nebo ztrátu informací o původním stavu chromatinu (Zouari et al., 2018).

V oblasti detekce chromozómových přestaveb je metoda Hi-C zcela zásadní, jelikož zobrazuje silné kontakty chromatinu na kontaktních mapách. Signál má nejvyšší hodnotu v rozpoznávacích bodech a postupně klesá se zvyšující se genetickou vzdáleností lokusů. Míra signálu na mapách napomáhá intuitivnímu rozpoznání chromozómových přestaveb. Konkrétně např. delece snižuje vzdálenost mezi segmenty genomu před a za rozpoznávacími body, což se zobrazuje jako nové chromatinové kontakty na rozpoznávacích bodech, se signály klesajícími nahoru od diagonály (Obr. 8A). Naproti tomu tandemová duplikace šíří nové chromatinové kontakty směrem dolů (Obr. 8B). Frekvence kontaktů se v případě invertované duplikace šíří na levou nebo pravou stranu v závislosti na směru amplifikace fragmentů (Obr. 8C). Při detekci reciprokových translokací a inverzí je pozorovatelný tzv. „butterfly shape“ (Obr. 8D), kdy se signál kontaktů šíří po obou stranách diagonály (K. Kim et al., 2022).

Metoda Hi-C přináší přesnější výsledky např. v kombinaci s technologiemi PacBio, které generují dlouhá čtení. Taková kombinace může pomoci přesnějšímu sestavování referenčních genomů *de novo* u nemodelových organismů. Takové genomy jsou potom vhodné pro analýzu konzervované syntenie druhů v rámci dané evoluční linie (Bai et al., 2021; Tian et al., 2021, 2022; Yekefenhazi et al., 2022).

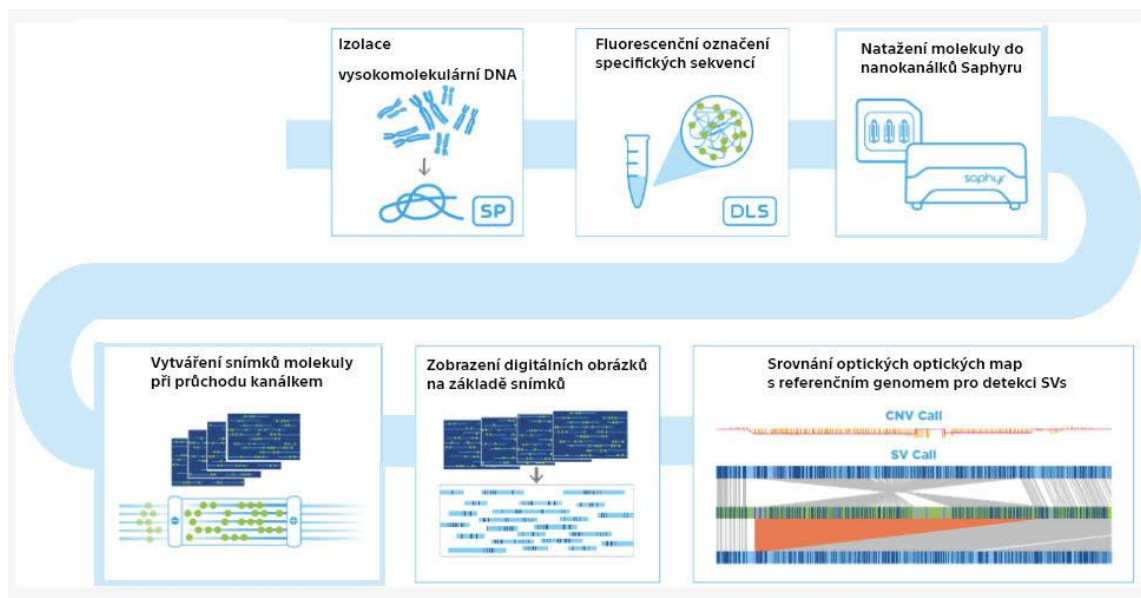
### 3.2.4 Optické mapování

Ani kombinace metod NGS a TGS nevede k sestavení genomu na chromozómové úrovni (Yuan et al., 2020). Pomoci může výše zmíněná metoda HiC, ale jinak v tomto ohledu na poli bioinformatiky zcela dominuje optické mapování genomu (neboli OGM, z angl. *optical genome mapping*), které dokáže analyzovat vysokomolekulární DNA (s průměrnou délkou fragmentů 400 kb). (Slatko et al., 2018; Yuan et al., 2020).

Přestože první protokol optického mapování byl navržen již roku 1993 (Das et al., 2010; Yuan et al., 2020), od roku 2012 přináší největší pokroky v její optimalizaci společnost Bionano Genomics. Principem metody je fluorescenční značení vybrané sekvence, obvykle hexamerního motivu, který se vyskytuje napříč genomem (u člověka a myši obvykle každých 5 kb) a toto značení vytvoří „čárový kód“ na linearizovaných molekulách DNA. Pro zajištění kvalitního vyhodnocování nasbíraných dat byly touto společností vyvinuty dvě platformy, Irys a Saphyr. Pro platformu Saphyr byl roku 2020 upraven protokol s názvem Direct Label and Stain (DLS), který byl optimalizován pro vysokou míru specifity (Obr. 9). Vysokomolekulární DNA je fluorescenčně značena v tomto případě enzymem Direct Label enzyme (DLE-1), který kovalentně navazuje fluorescenční značku na sekvencně specifická místa. Výhoda tohoto enzymu je, že nenarušuje okolní sekvence (internetový zdroj č. 3). Takto označená DNA je následně linearizována pomocí elektrického proudu a je natahována do nanokanáلكu. Při průchodu molekuly se pomocí fluorescenčního mikroskopu pořizují snímky jednotlivých oblastí molekuly, které jsou později sestaveny do optické mapy, jakmile jsou molekuly s podobným vzorem seskupeny do

klastrů. Zároveň jsou pomocí výpočetních programů určeny i vlastnosti jako je délka nebo hmotnost molekuly (Yuan et al., 2020; Jeffet et al., 2021).

Největší výhodou optického mapování je produkce dat, které nejsou zkreslené klonováním, amplifikací aj. celý proces vyžaduje nižší finanční i časovou investici laboratoří než klasické sekvenační metody (Bogas et al., 2017). Současně také poskytuje mapování chybějících sekvencí a doplňuje tak komplexní představu o genomu. Podobně jako v případě Hi-C, také optické mapování generuje nejlepší výsledky v kombinaci s jinými technologiemi, především TGS (Sun et al., 2020; Khan & Toledo, 2021).



**Obrázek 9:** Postup protokolu DLS, upraveno podle Sahajpal et al. (2021).

Strukturní varianty se poté určují při porovnávání optické mapy s referenčním genomem. Optickým mapováním je možné určit balancované i nebalancované SVs i CNVs (Lam et al., 2012; Mantere et al., 2021). Pomocí optického mapování bylo dosaženo nejlepších výsledků detekce strukturních variant v porovnání s jinými technologiemi (Teague et al., 2010; Lin et al., 2015).

Limity optického mapování tkví v tom, že poskytovaná data mají relativně nízké rozlišení, to se dá ale vyřešit kombinací s jinými technologiemi (Bogas et al., 2017). Dále současnými technologiemi není možné odhalovat balancované přestavby v (peri)centrických nebo telomerických oblastech, což zapříčiňuje především nedostatečné zmapování těchto oblastí v referenčních genomech (Dai et al., 2022).

### 3.3 Sestavování genomu

Jakmile jsou genomy osekvenovány, je dalším krokem jejich sestavení. V případě pozorování syntenních bloků a SVs je žádoucí, aby se dosáhlo úrovně, kdy je genom poskládaný na úroveň celých chromozómů (Damas et al., 2021). To vyžaduje kombinaci sekvenačních technologií s Hi-C a/nebo optickým mapováním (Burton et al., 2013). Myšlenka kombinace těchto technologií vedla k vytvoření

pipeliny CSA (z angl. *Chromosome-Scale Assembler*), která k sestavování genomu používá rozmanitá data generovaná různými technologiemi (Hi-C, 10x Genomics, HiFi PacBio, ONT aj.). Tato „pipeline“ umožnila automatizované sestavení genomu na úrovni chromozómů, čímž zvýšila rychlost i přesnost poskládání genomu (Kuhl et al., 2020). Největší bioinformatickou výzvu pro kvalitní sestavení genomu představují především chybně osekvenované repetitivní oblasti nebo např. chimérická čtení (fragmenty čtení se s referenčním genomem srovnávají na různých pozicích) (Nagarajan & Pop, 2013).

Rozvoj sekvenačních technologií napomohl umožnil i tzv. fázevení genomu, tzn. že jsme schopni genom poskládat jako otcovskou a mateřskou haploidní sadu, tj. sekvence chromozómů rodičů nejsou promíchané (Logsdon et al., 2021). K tomuto účelu se hojně využívá algoritmus zvaný trio binning, který k dosažení výsledků používá krátká čtení rodičovských genomů a dlouhá čtení genomu potomků (Koren et al., 2018). Další nadstavbou klasického sekvenování je také Mate-pair sekvenování, které umožňuje získat informace o vzdálenosti a orientaci dvou čtených sekvencí. Touto technikou pak lze docílit souvislejších dlouhých čtení a získání informací a struktury genomu, čímž se tento přístup stává ideálním nástrojem pro detekci SVs (Zeitouni et al., 2010).

Pro posouzení kvality sestavených genomů jsou určeny následující parametry: míra kontinuity, kompletnosti a správnosti. Automatizaci tohoto procesu napomohl vývoj algoritmu zvaný BUSCO (z angl. *Benchmarking Universal Single-Copy Orthologs*). Pracuje na základě identifikace univerzálních ortologů (tj. homologů genů vykonávajících stejnou funkci u blízce či vzdáleně příbuzných druhů), které jsou silně konzervované, jejich univerzální přítomnost v genomech organismů dělá z algoritmu BUSCO standardizovaný a objektivní nástroj k hodnocení kvality sestav genomů (Simão et al., 2015).

V posledních letech se stal hojně využívaným také algoritmus s názvem Hifiasm (z angl. *High fidelity assembler*), který pracuje s překryvem sekvencí z dlouhých čtení generovaných technologiemi třetí generace (OLC, z angl. *Overlapping layout consensus*). Hifiasm funguje na principu formování překrývajících se sekvencí do kontigů, na základě kterých následně vytvoří konsenzuální sekvence (Cheng et al., 2021). Hifiasm zahrnuje i korekci chyb, které produkují TGS. Dalším hojně používaným algoritmem je také Canu, který dokáže sestavit kompletní genomu na základě sekvencí generovaných PacBio nebo ONT. Stejně jako Hifiasm používá princip OLC, při čemž se odstraňují chyby ve čteních. V další fázi odhaluje mezery mezi sekvencemi mechanismem zvaným „gap filling“ (Koren et al., 2017).

### 3.3.1 Rekonstrukce genomu společných předků

Sledování rozdílů mezi druhy není možné, pokud nemáme s čím data srovnávat, proto je sestavení referenčních genomů absolutně zásadním krokem pro pozorování míry zachování nebo přeskupení syntenních bloků během evoluce. Aby bylo jasné, jaké události vedly k dnešní podobě karyotypů, je potřeba rekonstruovat genom společného předka. To se může provádět jak genomickými, tak cytogenomickými metodami. V dnešní době se již hojně využívají algoritmy, které vyhodnocují sekvence genomů s rozlišením jednoho nukleotidu (Damas et al., 2021). Většina algoritmů vyhodnocuje nejpravděpodobnější podobu chromozómů společného předka na základě parsimonie, v principu jde

tedy o to, že nejjednodušší cesta je pravděpodobně ta správná. Parsimonie předpokládá, že k současnému stavu se došlo nejmenším možným počtem změn a syntenní bloky příbuzných druhů představují stav společného předka (Damas et al., 2021). Např. ve studii z roku 2018 byl rekonstruován genom společného předka amniot, který byl sestaven na základě 61 osekenovaných genomů příbuzných druhů. Pomocí sledování změn na chromozómové úrovni je možné sledovat, jak se vyvíjely syntenní bloky a získávat informace o jejich vývoji. V tomto případě slouží fylogenetický strom jako důležitý nástroj pro posouzení a rekonstrukci vývoje těchto bloků (Sacerdot et al., 2018).

#### 3.3.1.1 *DESCHRAMBLER*

V důsledku rozvoje různých typů sekvenačních technologií byla sestavena řada genomů různé kvality a za použití různých postupů a algoritmů. DESCHRAMBLER je ovšem algoritmus, který umožňuje podrobné srovnávání genomů bez ohledu na jejich odlišnou kvalitu, což usnadňuje komparativní genomiku a v umožňuje hlubší vhled do strukturních přestaveb. Jeho klíčovým prvkem je schopnost rozpoznat struktury a porovnávat míru konzervované syntenie u souboru genomů a po identifikaci a porovnání míry konzervované syntenie vytváří rekonstrukci zkoumaných přestaveb, čímž můžeme získat představu o ancestrální podobě karyotypu a zkoumat, jaké evoluční události vedly k jeho současnému uspořádání. Algoritmus byl poprvé představen roku 2017, kdy byl srovnáván s metodou FISH. Nejen, že se výsledky téměř shodovaly, ale DECSHRAMBLER detekoval i přestavby, které předchozí studie nezaznamenaly (J. Kim et al., 2017).

### 3.4 Algoritmy pro detekci SVs

Sekvenační technologie produkují obrovské množství dat, které je ale potřeba dále zpracovávat. Pro zpracování bylo vyvinuto několik algoritmů, které se rozdělují na zarovnávací (tzv. aligners) a vyvolávací (tzv. callers). Zarovnávací algoritmy mapují sekvenovaný genom na referenční, callers na ně dále navazují, a identifikují konkrétní SVs (Mahmoud et al., 2019).

#### 3.4.1 Alligners

Algoritmů, které provádějí zarovnání genomů ze sekvenačních map je velké množství a stále se vyvíjejí další, např. BLAST, MULTIZ, Mishmap2 aj. Přičemž různé algoritmy zároveň generují různé výsledky. V současné době nejpoužívanějším algoritmem je Cactus, který pro vyhodnocení dat nepotřebuje referenční genom a výsledky tak nejsou zkresleny. Jeho schopnost pracovat s dlouhými čtecími sekvencemi a analyzovat strukturní varianty mu umožňuje přesně sestavit a porovnávat genomy, díky čemuž se snadněji detekují SVs. (Armstrong et al., 2020). Nevýhodou algoritmu Cactus je však to, že v tuto chvíli dokáže zpracovat pouze několik stovek genomů (Damas et al., 2021). Aby však bylo dosaženo komplexní databanky pro porovnávání genomů, je potřeba do budoucna vytvářet algoritmy, které budou umět sesadit velké množství genomových dat najednou, což je důležité vzhledem k tomu,

jak obrovské množství dat produkují sekvenační technologie. Dalším často používaným algoritmem je Minimap2, který oproti algoritmu Cactus využívá pro analýzu výsledků referenčního genomu, identifikuje tak podobnosti a rozdíly variant mezi analyzovanými genomovými sekvencemi a referenčním genomem. Je založen na principu aligneru, který se snaží dosáhnout optimálního zarovnání mezi sekvencemi na základě podobnosti sekvencí a pravděpodobnosti jejich variant. Analyzuje především sekvenační data generovaná technologiemi třetí generace. Je schopen mapovat dlouhé sekvence s vysokou rychlostí a efektivitou, je 3-4x rychlejší než jiné alignery (H. Li, 2018). Minimap2 i Cactus jsou volně přístupné (open source) algoritmy, jejich zdrojové kódy jsou využívány pro analýzu genomů velkého rozsahu.

### 3.4.2 Callers

Po fázi sestavení a zarovnání vůči referenčnímu genomu je potřeba detekce a identifikace jednotlivých SVs. Pomocí srovnávání s referenčním genomem jsou identifikovány rozdíly, které mohou představovat strukturní změnu. Pro detekci SVs se používají čtyři základní přístupy (signály): Sestavení genomu (neboli AS, z angl. *Assembly*), Rozdělené čtení (neboli SR, z angl. *Split reads*), Hloubka čtení (neboli RD, z angl. *Read depth*) a Páry čtení (neboli RP, z angl. *Read pair*).

Jak již může napovědět název, přístupy AS jsou založeny na sestavení genomu. Konkrétně jde o sestavení krátkých sekvencí produkovaných NGS do delších kontigů. S referenčním genomem se srovnávají právě tyto kontigy, protože jsou delší než sekvenované fragmenty, což umožňuje snazší lokalizaci SVs. Přístup SR mapuje již osekvenované molekuly DNA a porovnává je s referenčními genomy. Pomocí zarovnávání tak získává informace o možnostech výskytu SVs. Signály RP jsou nejpoužívanějším přístupem pro detekci SVs, především z toho důvodu, že identifikují většinu typů přestaveb. Metoda porovnává páry čtení na základě prahových hodnot, jinými slovy počítá délky čtení mezi dvěma konci a dále je porovnává s referenčním genomem. Posledním přístupem je RD, který umí detekovat pouze delece a duplikace, přičemž principem je využití mechanismu hloubky čtení, pomocí kterého RD porovnává namapované sekvence s referenčním genomem. Jakmile detekuje změny v hloubce pokrytí, může se v tomto místě nacházet přestavba (Kathuria & Ratan, 2020; Lin et al., 2015; Sahlin et al., 2017). Signály SR a RP využívá např. algoritmus Delly (Kuzniar et al., 2020). Nejpoužívanějším algoritmem, který vyhodnocuje osekvenované genomy na základě výše pospaných přístupů, byl po dlouhou dobu BreakDancer, který používá RP (Chen et al., 2009). Vzhledem k rozmanitosti SVs se nedá v současnosti ani jeden algoritmus považovat za ideální, proto je dobré kombinovat různé typy dohromady (Lin et al., 2015; Ho et al., 2020; Kuzniar et al., 2020; Yang, 2020). Ke každé metodě NGS jsou navíc vyvíjeny algoritmy šité na míru datům, které produkují, např. pro SMRT od PacBio se používá algoritmus zvaný SMRT link nebo SMRT-SV (Chaisson et al., 2015). Pro ONT je to např. EPI2ME (Hatfield et al., 2020) nebo NanoSV (Cretu Stancu et al., 2017).

### 3.5 Bioinformatické pipeliney

Jak již bylo zmíněno, detekce strukturních variant je komplexní proces, který vyžaduje překonání mnoha bioinformatických překážek. Pro optimalizaci a urychlení procesu se proto vytváří tzv. pipeliney, které se skládají z několika po sobě jdoucích kroků a předávají si data mezi sebou. Prvním krokem je vždy sekvenace dat a jejich následné zpracování, tedy sestavení genomu. Dalším krokem zpravidla bývá mapování na referenční genom pomocí aligners, čímž se získají potřebné informace o jednotlivých sekvencích a jejich pozici v genomu. V dalším kroku přichází na řadu samotná detekce SVs, jejíž výsledky je poté potřeba podrobit validizaci, při které jsou používány např. cytogenomické metody. Jako poslední přichází na řadu anotace dat, během níž se SVs přiřazují konkrétní biologické vlastnosti a vliv na genom. To, jaké technologie a algoritmy se používají v jednotlivých pipelinech, se velmi liší (příklady studií: Lin et al., 2015; Xia et al., 2018; Zhou et al., 2022).

Jedním z největších projektů, který využívá bioinformatické pipeliney pro osekvenování genomů a jejich následné srovnávání na chromozómové úrovni, je Vertebrate Genome Project (VGP). Ve své původní verzi byly genomy nejprve sekvenovány pomocí PacBio SMRT a následně sestavovány na úrovni chromozómů pomocí optického mapování (Bionano) a Hi-C technologií. Z důvodů popsaných v kapitolách výše bylo v další verzi PacBio SMRT vyměněno za PacBio HiFi, což přináší přesnější data s menší chybovostí (internetový zdroj č.4).

## 4 Závěr

Chromozómové přestavby a jejich polymorfizmy hrají důležitou úlohu ve speciaci, diverzifikaci druhů a adaptaci na podmínky okolního prostředí. Takové strukturní změny mohou ovlivňovat funkci genů, regulaci genové exprese, četnost a distribuci rekombinace a další důležité parametry, které souvisejí s funkční dynamikou genomu a jeho evolucí. I přesto, že téma chromozómových přestaveb je v současné době hojně zkoumáno, je obvykle obtížné dokázat souvislost mezi vznikem přestavby a fenotypovým projevem. Po dlouhou dobu byly pro zkoumání přestaveb používány cytogenomické a cytogenetické metody, jejichž rozlišovací schopnost však neumožňuje odhalit celkové spektrum přestaveb v daném genomu. V posledních letech se však uskutečnily značné pokroky na poli genomiky a bioinformatiky, které umožňují detailnější vhled do této problematiky.

Nejen prokázání kauzality, ale také samotná detekce CRs a SVs je sama o sobě výzvou. Z výpočetního hlediska je největším problémem při detekci SVs vysoká míra falešně pozitivních výsledků, kterou současné technologie generují. I z tohoto důvodu je pro správnou detekci a identifikaci důležitý kvalitně poskládaný referenční genom, či jiné srovnávací testy, podle kterých by se dala ověřovat validita výsledků (Lin et al., 2015). Sekvenační technologie se podle typu templátu pro sekvenování, délky čtení, chybovosti a období jejich zavedení do provozu dají rozdělit na NGS s TGS. Sekvenačními technologiím druhé generace se nedá odeprít jejich důležitost při sekvenování a analýze velkého množství genomů, i přesto jsou ale pro studium CRs s SVs nevhodné, vzhledem k tomu že jejich krátká sekvenční čtení nejsou schopná překlenout, a tudíž správně poskládat dlouhé úseky repetitivních sekvencí (Chaisson et al., 2015; Ramos & Antunes, 2022). Na druhou stranu, technologie NGS (resp. Illumina) nacházejí díky své vysoké přesnosti uplatnění při korekci chyb vzniklých při sekvenování metodami TGS (Mahmoud et al., 2019). Pro opravu sekvenačních chyb bylo vyvinuto několik algoritmů, z nichž nejpoužívanější (podrobněji popsany výše) je Canu (Koren et al., 2017). Zajímavé je, že i když jsou pro detekci SVs vhodnější metody TGS, většina přídatných algoritmů je vyvíjená pro metody NGS, přičemž pro TGS nejsou použitelné (Sedlazeck et al., 2018a). Nejlepší sekvenační data generuje přístup HiFi od PacBio a Nanopore sequencing od ONT (délka čtení: 10 – 1000 kb). Při srovnávání těchto dvou technologií bylo prokázáno, že ONT a jeho ULR (ultradelouhé ready) jsou ideální pro překonání oblastí dlouhých úseků repetitivní DNA, ale generují poměrně velké množství chyb. Naproti tomu HiFi PacBio nevytváří tak dlouhá sekvenční čtení, ale jejich přesnost je až 99,9 % (Lang et al., 2020). Nelze však opomenout fakt, že cytogenomické metody se stále používají a slouží jako doplnění sekvenačních metod (O'Connor et al., 2017; Deakin et al., 2019; Lee et al., 2020), v některých případech dokonce jako parametr, dle kterého se určuje validita výsledků sekvenačních technologií (Lee et al., 2020; Rhie et al., 2021). Mohou přinést informaci o počtu vazebných skupin a uspořádání obtížně sekvenovatelných úseků heterochromatinu/repetitivních sekvencí. Zároveň mohou i snížit komplexitu sekvenování, pokud se s jejich pomocí separátně sekvenují jednotlivé chromozómy (Doležel et al., 2014; Iannucci et al. 2021). Dalším přínosem jsou technologie Hi-C, optické mapování a

10x Genomics, které umožňují poskládat genom na úroveň celých chromozómů (chromosome-level assembly), ideálně i bez chybějících na repetice bohatých (sub)telomerických oblastí (T2T assembly). Tyto technologie jsou ideálním nástrojem pro sestavování genomu *de novo*, neboť ze své podstaty přímo generují data s informacemi o kontaktech mezi chromozómy a chromatinem.

Obrovské množství dat, které produkují TGS metody, jsou dále zpracovávány. Vznikají tím kvalitně poskládané genomy, které se následně zarovnávají vůči referenčnímu genomu. Jakmile je k dispozici kvalitní genomová sestava, může dojít na detekci konkrétních SVs. To se provádí nástroji zvanými callers, které jsou optimalizovány pro konkrétní technologie (viz Tab. 1, sloupec algoritmy).

Obecně se dá tedy říci, že pro detekci a identifikaci komplexních chromozómových přestaveb jsou ideální sekvenační technologie třetí generace. Problém je pouze ten, že se vyznačují vysokou chybovostí, a proto je vhodné jejich výsledky doplňovat jak o krátká sekvenční čtení (Illumina), tak i o klasické cytogenomické a cytogenetické metody. V Tab. 1 jsou shrnuty současné genomické technologie, jejich stručná charakteristika, výhody a nevýhody a také jaké algoritmy jsou pro ně vyvinuty pro vyhodnocování SVs. Pravděpodobně nejlepším řešením pokrytí všech kritérií pro správnou identifikaci přestavby je však kombinace různých metod. I z toho důvodu se v současné době na trhu uchytily pipeline, které používají několik metod dohromady, sdílí si mezi sebou data a poté je společně vyhodnocují (Abel et al., 2020).

Soudě podle ohromného množství článků, které se podstatou vzniku a funkcí CRs a SVs (v různých kontextech) zabývá se jedná se o velmi aktuální téma a aktivní pole výzkumu, které má velkou perspektivu do budoucna. Tyto poznatky nám mohou pomoci nejen lépe pochopit evoluci druhů a reprodukčních překážek mezi nimi, ale také mechanismy souvisejícími s vrozenými vadami a nádorovou transformací. Mohou mít také aplikační potenciál z hlediska konzervační genetiky, ekologie a zemědělství. Je proto žádoucí, aby byly vyvinuty metody a protokoly aplikovatelné na celé spektrum eukaryotických organizmů (Sedlazeck et al., 2018a).

Metoda	Charakteristika	Výhody	Nevýhody	Algoritmy
<b>Illumina</b>	Krátká čtení (v řádech stovek bp) Paralelní sekvenování	Nízká chybovost Nízká cena	Krátká čtení, problém především u repetitivních sekvencí Amplifikační bias	Minimap2 LongRanger GROC-SVs
<b>SMRT PacBio</b>	Dlouhá čtení (10 kb – 100 kb) Používá se především k sestavování genomu <i>de novo</i> , detekci SVs a epigenetických modifikací	Produkuje dlouhá čtení Žádný amplifikační bias	Vysoká míra chybovosti, až 15 % Vysoká cena oproti NGS	Minimap2 SMRT-SV SMRT link MARVEL
<b>HiFi PacBio</b>	Délka sekvencí vyšší než 10 kb Sestavování genomu <i>de novo</i> Detekce SVs	Nízká chybovost, přesnost až 99,5 %	Vysoká cena	GATK SMRT link DeepVariant
<b>ONT</b>	Dlouhá čtení (10 kb – více než 1 Mb) Používá se k sestavování genomu <i>de novo</i> , detekci rozsáhlých SVs	Velmi dlouhá čtení, umí překonat repetitivní oblasti	Vysoká cena Vysoká chybovost	NanoSV EPI2ME
<b>Hi-C</b>	Vytváří páry krátkých čtení (menší než 100 bp), které zobrazují kontakty v chromatinu Podpůrná metoda pro sestavení genomu na úrovni chromozómů	Zobrazení interakcí chromatinu	Vysoká cena Nemusí vždy odhalit všechny interakce na velké vzdálenosti	Hic- breakfinder HiCnv HiCtrans
<b>BioNano optické mapování</b>	Mapování dlouhých molekul DNA (250 kb a delší) pomocí fluorescenčních sond Sestavení genomu na úrovni chromozómů	Nízká cena Žádný amplifikační bias	Obtížný vývoj vhodných algoritmů pro detekci SVs	BioNano Acces

**Tabulka 1:** Tabulka současně používaných genomických metod a jejich algoritmů pro detekci SVs. Sestaveno na základě prací Sedlazeck et al., (2018a) a Mahmoud et al., (2019).

## Literatura

- Abdullaev, E. T., Umarova, I. R., & Arndt, P. F. (2021). Modelling segmental duplications in the human genome. *BMC Genomics*, 22(1), 496.
- Abel, H. J., Larson, D. E., Regier, A. A., Chiang, C., Das, I., Kanchi, K. L., et al. (2020). Mapping and characterization of structural variation in 17,795 human genomes. *Nature*, 583(7814), 83–89.
- Alexeev, N., & Alekseyev, M. A. (2017). Estimation of the true evolutionary distance under the fragile breakage model. *BMC Genomics*, 18(S4), 356.
- Alkan, C., Sajjadian, S., & Eichler, E. E. (2011). Limitations of next-generation genome sequence assembly. *Nature Methods*, 8(1), 61–65.
- Álvarez-González, L., Burden, F., Doddamani, D., Malinverni, R., Leach, E., Marín-García, C et al. (2022). 3D chromatin remodelling in the germ line modulates genome evolutionary plasticity. *Nat Commun*, 13(1), 2608.
- Anand, R. P., Tsaponina, O., Greenwell, P. W., Lee, C.-S., Du, W., Petes, T. D. et al. (2014). Chromosome rearrangements via template switching between diverged repeated sequences. *Genes Dev*, 28(21), 2394–2406.
- Armstrong, J., Hickey, G., Diekhans, M., Fiddes, I. T., Novak, A. M., Deran, A., et al. (2020). Progressive Cactus is a multiple-genome aligner for the thousand-genome era. *Nature*, 587(7833), 246–251.
- B. Zouari, Y., Molitor, A. M., & Sexton, T. (2018). Sailing the Hi-C's. In *Nuclear Architecture and Dynamics* (pp. 457–473). Elsevier.
- Bai, Y., Gong, J., Zhou, Z., Li, B., Zhao, J., Ke, Q. et al. (2021). Chromosome-Level Assembly of the Southern Rock Bream (*Oplegnathus fasciatus*) Genome Using PacBio and Hi-C Technologies. *Front Genet*, 12.
- Balachandran, P., Walawalkar, I. A., Flores, J. I., Dayton, J. N., Audano, P. A., & Beck, C. R. (2022). Transposable element-mediated rearrangements are prevalent in human genomes. *Nat Commun*, 13(1), 7115.
- Bandyopadhyay, R., Heller, A., Knox-DuBois, C., McCaskill, C., Berend, S. A., Page, S. L., et al. (2002). Parental Origin and Timing of De Novo Robertsonian Translocation Formation. *Am J Hum Genet*, 71(6), 1456–62.
- Barra, V., & Fachinetti, D. (2018). The dark side of centromeres: types, causes and consequences of structural abnormalities implicating centromeric DNA. *Nat Commun*, 9(1), 4340.
- Belton, J.-M., McCord, R. P., Gibcus, J. H., Naumova, N., Zhan, Y., & Dekker, J. (2012). Hi-C: A comprehensive technique to capture the conformation of genomes. *Methods*, 58(3), 268–276.
- Berthelot, C., Muffato, M., Abecassis, J., & Roest Crollius, H. (2015). The 3D Organization of Chromatin Explains Evolutionary Fragile Genomic Regions. *Cell Rep*, 10(11), 1913–1924.
- Blackmon, H., Justison, J., Mayrose, I., & Goldberg, E. E. (2019). Meiotic drive shapes rates of karyotype evolution in mammals. *Evolution*, 73(3), 511–523.
- Bogas, D., Nyberg, L., Pacheco, R., Azevedo, N. F., Beech, J. P., Gomila, M., et al. (2017). Applications of optical DNA mapping in microbiology. *BioTechniques*, 62(6), 255–267.
- Branton, D., Deamer, D. W., Marziali, A., Bayley, H., Benner, S. A., Butler, T., et al. (2008). The potential and challenges of nanopore sequencing. *Nat Biotechnol*, 26(10), 1146–53.
- Brown, J. D., & O'Neill, R. J. (2010). Chromosomes, Conflict, and Epigenetics: Chromosomal Speciation Revisited. *Annu Rev Genomics Hum Genet*, 11(1), 291–316.
- Burton, J. N., Adey, A., Patwardhan, R. P., Qiu, R., Kitzman, J. O., & Shendure, J. (2013). Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nat Biotechnol*, 31(12), 1119–25.
- Carneiro, M. O., Russ, C., Ross, M. G., Gabriel, S. B., Nusbaum, C., & DePristo, M. A. (2012). Pacific biosciences sequencing technology for genotyping and variation discovery in human data. *BMC Genomics*, 13(1), 375.

- Chaisson, M. J. P., Huddleston, J., Dennis, M. Y., Sudmant, P. H., Malig, M., Hormozdiari, F., et al. (2015).** Resolving the complexity of the human genome using single-molecule sequencing. *Nature*, *517*(7536), 608–11.
- Chen, K., Wallis, J. W., McLellan, M. D., Larson, D. E., Kalicki, J. M., Pohl, C. S., et al. (2009).** BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. *Nat Methods*, *6*(9), 677–81.
- Cheng, H., Concepcion, G. T., Feng, X., Zhang, H., & Li, H. (2021).** Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods*, *18*(2), 170–75.
- Chi Y, Shi J, Xing D and Tan L (2022),** Every gene everywhere all at once: High-precision measurement of 3D chromosome architecture with singlecell Hi-C. *Front. Mol. Biosci.* *9*:959688.
- Chiang, C., Scott, A. J., Davis, J. R., Tsang, E. K., Li, X., Kim, Y., et al. (2017).** The impact of structural variation on human gene expression. *Nat Gen*, *49*(5), 692–99.
- Clark, F. E., & Akera, T. (2021).** Unravelling the mystery of female meiotic drive: where we are. *Open Biology*, *11*(9).
- Coughlan, J. M., & Matute, D. R. (2020).** The importance of intrinsic postzygotic barriers throughout the speciation process. *Philos Trans R Soc B Biol Sc*, *375*(1806), 20190533.
- Cremer, T., & Cremer, M. (2010).** Chromosome Territories. *Cold Spring Harb Perspect Biol*, *2*(3), a003889.
- Cretu Stancu, M., van Roosmalen, M. J., Renkens, I., Nieboer, M. M., Middelkamp, S., de Ligt, J., et al. (2017).** Mapping and phasing of structural variation in patient genomes using nanopore sequencing. *Nat Commun*, *8*(1), 1326.
- Dai, P., Zhu, X., Pei, Y., Chen, P., Li, J., Gao, Z., et al. (2022).** Evaluation of optical genome mapping for detecting chromosomal translocation in clinical cytogenetics. *Mol Genet Genomic Med* *10*(6).
- Damas, J., Corbo, M., & Lewin, H. A. (2021).** Vertebrate Chromosome Evolution. *Annu Rev Anim Biosci*, *9*(1), 1–27.
- Damas, J., O’Connor, R., Farré, M., Lenis, V. P. E., Martell, H. J., Mandawala, A., et al. (2017).** Upgrading short-read animal genome assemblies to chromosome level using comparative genomics and a universal probe set. *Genome Res*, *27*(5), 875–84.
- Das, S. K., Austin, M. D., Akana, M. C., Deshpande, P., Cao, H., & Xiao, M. (2010).** Single molecule linear analysis of DNA in nano-channel labeled with sequence specific fluorescent probes. *Nucleic Acids Res*, *38*(18), e177.
- De Coster, W., De Rijk, P., De Roeck, A., De Pooter, T., D’Hert, S., Strazisar, M., et al. (2019).** Structural variants identified by Oxford Nanopore PromethION sequencing of the human genome. *Genome Res*, *29*(7), 1178–87.
- Pardo-Manuel de Villena, F. P.-M., & Sapienza, C. (2001).** Female Meiosis Drives Karyotypic Evolution in Mammals. *Genetics*, *159*(3), 1179–89.
- Deakin, J. E., Potter, S., O’Neill, R., Ruiz-Herrera, A., Cioffi, M. B., Eldridge, M. D. B., et al. (2019).** Chromosomics: Bridging the Gap between Genomes and Chromosomes. *Genes*, *10*(8), 627.
- Delprat, A., Negre, B., Puig, M., & Ruiz, A. (2009).** The Transposon Galileo Generates Natural Chromosomal Inversions in *Drosophila* by Ectopic Recombination. *PLoS ONE*, *4*(11), e7883.
- Dickinson, G. R., Lee, D. J., & Wallace, H. M. (2012).** The influence of pre- and post-zygotic barriers on interspecific *Corymbia* hybridization. *Ann Bot*, *109*(7), 1215–26.
- Dixkens, C., Klett, C., Bruch, J., Kollak, A., Serov, O. L., Zhdanova, N., et al. (1998).** ZOO-FISH analysis in insectivores: “Evolution extols the virtue of the status quo.” *Cytogenet Genome Res* *80*(1–4), 61–7.
- Dobigny, G., Britton-Davidian, J., & Robinson, T. J. (2017).** Chromosomal polymorphism in mammals: an evolutionary perspective. *Biological Reviews*, *92*(1), 1–21.
- Dobigny, G., Ducroz, J.-F., Robinson, T. J., & Volobouev, vityal. (2004).** Cytogenetics and Cladistics. *Syst Biol*, *53*(3), 470–84.

- Dodsworth, S., Chase, M. W., & Leitch, A. R. (2016).** Is post-polyploidization diploidization the key to the evolutionary success of angiosperms? *Bot J Linn Soc*, *180*(1), 1–5.
- Doležel, J., Vrána, J., Cápál, P., Kubaláková, M., Burešová, V., & Šimková, H. (2014).** Advances in plant chromosome genomics. *Biotechnol Advances*, *32*(1), 122–136.
- Dunning, M. J., Barbosa-Morais, N. L., Lynch, A. G., Tavaré, S., & Ritchie, M. E. (2008).** Statistical issues in the analysis of Illumina data. *BMC Bioinformatics*, *9*(1), 85.
- Dutta, U. R., Bhattacharjee, A., Bahal, A., Posaapally, L. P., Lone, K. A., Bathula, S., & Dalal, A. (2022).** Cytogenomic Characterization of a Novel de novo Balanced Reciprocal Translocation t(1;12) by Genome Sequencing Leading to Fusion Gene Formation of EYA3/EFCAB4b; *Mol Syndromol*, *13*(5), 370–80.
- Ehrlich, J., Sankoff, D., & Nadeau, J. H. (1997).** Synteny Conservation and Chromosome Rearrangements During Mammalian Evolution. *Genetics*, *147*(1), 289–96.
- Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., et al. (2009).** Real-Time DNA Sequencing from Single Polymerase Molecules. *Science*, *323*(5910), 133–138.
- English, A. C., Salerno, W. J., Hampton, O. A., Gonzaga-Jauregui, C., Ambreth, S., Ritter, D. I., Beck, C. R., et al. (2015).** Assessing structural variation in a personal genome—towards a human reference diploid genome. *BMC Genomics*, *16*(1), 286.
- Escalante, A. E., Jardón Barbolla, L., Ramírez-Barahona, S., & Eguiarte, L. E. (2014).** The study of biodiversity in the era of massive sequencing. *Rev Mex Biodivers*, *85*(4), 1249–64.
- Fan, H., Wu, Q., Wei, F., Yang, F., Ng, B. L., & Hu, Y. (2019).** Chromosome-level genome assembly for giant panda provides novel insights into Carnivora chromosome evolution. *Genome Biology*, *20*(1), 267.
- Faria, R., & Navarro, A. (2010).** Chromosomal speciation revisited: rearranging theory with pieces of evidence. *Trends Ecol Evol*, *25*(11), 660–9.
- Ferguson-Smith, M. A., & Trifonov, V. (2007).** Mammalian karyotype evolution. *Nat Rev Genet*, *8*(12), 950–62.
- Feuk, L., Carson, A. R., & Scherer, S. W. (2006).** Structural variation in the human genome. *Nat Rev Genet*, *7*(2), 85–97.
- Gao, L., Xu, W., Xin, T., & Song, J. (2023).** Application of third-generation sequencing to herbal genomics. *Front Plant Sci*, *14*.
- Garagna, S., Page, J., Fernandez-Donoso, R., Zuccotti, M., & Searle, J. B. (2014).** The Robertsonian phenomenon in the house mouse: mutation, meiosis and speciation. *Chromosoma*, *123*(6), 529–44.
- Goel, V. Y., & Hansen, A. S. (2021).** The macro and micro of chromosome conformation capture. *Wiley Interdiscip Rev Dev Biol*, *10*(6), e395.
- Goettel, W., & Messing, J. (2010).** Divergence of gene regulation through chromosomal rearrangements. *BMC Genomics*, *11*(1), 678.
- Gridina, M., Mozheiko, E., Valeev, E., Nazarenko, L. P., Lopatkina, M. E., Markova, Z. G., et al. (2021).** A cookbook for DNase Hi-C. *Epigenetics Chromatin*, *14*(1), 15.
- Guarracino, A., Buonaiuto, S., de Lima, L. G., Potapova, T., Rhie, A., Koren, S., et al. (2023).** Recombination between heterologous human acrocentric chromosomes. *Nature*, *617*(7960), 335–43.
- Gutiérrez-Valencia, J., Hughes, P. W., Berdan, E. L., & Slotte, T. (2021).** The Genomic Architecture and Evolutionary Fates of Supergenes. *Genome Biol Evol*, *13*(5).
- Harewood, L., & Fraser, P. (2014).** The impact of chromosomal rearrangements on regulation of gene expression. *Hum Mol Genet*, *23*(R1), R76–82.
- Hatfield, R. G., Batista, F. M., Bean, T. P., Fonseca, V. G., Santos, A., Turner, A. D., Lewis, A., et al. (2020).** The Application of Nanopore Sequencing Technology to the Study of Dinoflagellates: A Proof of Concept Study for Rapid Sequence-Based Discrimination of Potentially Harmful Algae. *Front Microbiol*, *11*.

- Heslop-Harrison, J. S. (Pat), Schwarzacher, T., & Liu, Q. (2023).** Polyploidy: its consequences and enabling role in plant diversification and evolution. *Ann Bot*, 131(1), 1–10.
- Ho, S. S., Urban, A. E., & Mills, R. E. (2020).** Structural variation in the sequencing era. *Nat Rev Genet*, 21(3), 171–89.
- Hu, Q., Maurais, E. G., & Ly, P. (2020).** Cellular and genomic approaches for exploring structural chromosomal rearrangements. *Chromosome Res*, 28(1), 19–30.
- Huber, D., Voith von Voithenberg, L., & Kaigala, G. V. (2018).** Fluorescence in situ hybridization (FISH): History, limitations and what to expect from micro-scale FISH? *Micro Nano Engineering*, 1, 15–24.
- Iannucci, A., Makunin, A. I., Lisachov, A. P., Ciofi, C., Stanyon, R., Svartman, M., et al. (2021).** Bridging the Gap between Vertebrate Cytogenetics and Genomics with Single-Chromosome Sequencing (ChromSeq). *Genes*, 12(1), 124.
- Jackson, B. C. (2011).** Recombination-suppression: how many mechanisms for chromosomal speciation? *Genetica*, 139(3), 393–402.
- Jain, M., Koren, S., Miga, K. H., Quick, J., Rand, A. C., Sasani, T. A., et al. (2018).** Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat Biotechnol*, 36(4), 338–45.
- Jarmuz-Szymczak, M., Janiszewska, J., Szyfter, K., & Shaffer, L. G. (2014).** Narrowing the localization of the region breakpoint in most frequent Robertsonian translocations. *Chromosome Res*, 22(4), 517–32.
- Jeffet, J., Margalit, S., Michaeli, Y., & Ebenstein, Y. (2021).** Single-molecule optical genome mapping in nanochannels: multidisciplinary at the nanoscale. *Essays Biochem*, 65(1), 51–66.
- Jiang, J. (2019).** Fluorescence in situ hybridization in plants: recent developments and future applications. *Chromosome Res*, 27(3), 153–65.
- Kasianowicz, J. J., Brandin, E., Branton, D., & Deamer, D. W. (1996).** Characterization of individual polynucleotide molecules using a membrane channel. *Proc Natl Acad Sci U S A*, 93(24), 13770–3.
- Khan, W. A., & Toledo, D. M. (2021).** Applications of Optical Genome Mapping in Next-Generation Cytogenetics and Genomics. *Advances in Molecular Pathology*, 4, 27–36.
- Kim, J., Farré, M., Auvil, L., Capitanu, B., Larkin, D. M., Ma, J., & Lewin, H. A. (2017).** Reconstruction and evolutionary history of eutherian chromosomes. *Proc Natl Acad Sci U S A*, 114(27).
- Kim, K., Kim, M., Kim, Y., Lee, D., & Jung, I. (2022).** Hi-C as a molecular rangefinder to examine genomic rearrangements. *Semin Cell Dev Biol* 121, 161–70.
- Kono, N., & Arakawa, K. (2019).** Nanopore sequencing: Review of potential applications in functional genomics. *Dev Growth Differ*, 61(5), 316–26.
- Koren, S., Rhie, A., Walenz, B. P., Dilthey, A. T., Bickhart, D. M., Kingan, S. B., et al. (2018).** De novo assembly of haplotype-resolved genomes with trio binning. *Nat Biotechnol*, 36(12), 1174–82.
- Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., & Phillippy, A. M. (2017).** Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Res*, 27(5), 722–36.
- Kuhl, H., Li, L., Wuertz, S., Stöck, M., Liang, X.-F., & Klopp, C. (2020).** CSA: A high-throughput chromosome-scale assembly pipeline for vertebrate genomes. *GigaScience*, 9(5).
- Kumon, T., & Lampson, M. A. (2022).** Evolution of eukaryotic centromeres by drive and suppression of selfish genetic elements. *Semin Cell Dev Biol*, 128, 51–60.
- Kuzniar, A., Maassen, J., Verhoeven, S., Santuari, L., Shneider, C., Kloosterman, W. P., et al. (2020).** sv-callers: a highly portable parallel workflow for structural variant detection in whole-genome sequence data. *PeerJ*, 8, e8214.

- Lam, E., Hastie, A., Lin, C., Ehrlich, D., Das, S., Austin, M., et al. (2012). Genome mapping on nanochannel arrays for structural variation analysis and sequence assembly. *Nat Biotechnol*, 30(8), 771–6.
- Lang, D., Zhang, S., Ren, P., Liang, F., Sun, Z., Meng, G., et al. (2020). Comparison of the two up-to-date sequencing technologies for genome assembly: HiFi reads of Pacific Biosciences Sequel II system and ultralong reads of Oxford Nanopore. *GigaScience*, 9(12).
- Larkin, D. M., Pape, G., Donthu, R., Auvil, L., Welge, M., & Lewin, H. A. (2009). Breakpoint regions and homologous synteny blocks in chromosomes have different evolutionary histories. *Genome Res*, 19(5), 770–7.
- Lee, L. S., Navarro-Domínguez, B. M., Wu, Z., Montiel, E. E., Badenhorst, D., Bista, B., et al. (2020). Karyotypic Evolution of Sauropsid Vertebrates Illuminated by Optical and Physical Mapping of the Painted Turtle and Slider Turtle Genomes. *Genes*, 11(8), 928.
- Leffak, M. (2017). Break-induced replication links microsatellite expansion to complex genome rearrangements. *BioEssays*, 39(8), 1700025.
- Levan, A., Fredga, K., & Sandberg, A. A. (2009). Nomenclature For Centromeric Position On Cchromosomes. *Hereditas*, 52(2), 201–20.
- Levene, M. J., Korlach, J., Turner, S. W., Foquet, M., Craighead, H. G., & Webb, W. W. (2003). Zero-Mode Waveguides for Single-Molecule Analysis at High Concentrations. *Science*, 299(5607), 682–6.
- Levsky, J. M., & Singer, R. H. (2003). Fluorescence in situ hybridization: past, present and future. *J Cell Sci*, 116(14), 2833–8.
- Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, 34(18), 3094–100.
- Li, J., Gao, L., & Ye, Y. (2023). HiSV: A control-free method for structural variation detection from Hi-C data. *PLOS Computat Biol*, 19(1), e1010760.
- Li, S.-F., Su, T., Cheng, G.-Q., Wang, B.-X., Li, X., Deng, C.-L., et al. (2017). Chromosome Evolution in Connection with Repetitive Sequences and Epigenetics in Plants. *Genes*, 8(10), 290.
- Li, X., Ellis, E., Plotkin, D., Imada, Y., Yago, M., Heckenhauer, J., et al. (2021). First Annotated Genome of a Mandibulate Moth, *Neomicropteryx cornuta*, Generated Using PacBio HiFi Sequencing. *Genome Biol Evol*, 13(10).
- Li, X., Wang, J., Yu, Y., Li, G., Wang, J., Li, C., et al. (2023). Genomic rearrangements and evolutionary changes in 3D chromatin topologies in the cotton tribe (Gossypieae). *BMC Biology*, 21(1), 56.
- Liao, Y., Liu, Z., Zhang, Y., Lu, P., Wen, L., & Tang, F. (2023). High-throughput and high-sensitivity full-length single-cell RNA-seq analysis on third-generation sequencing platform. *Cell Discovery*, 9(1), 5.
- Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragooczy, T., Telling, A., et al. (2009). Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science*, 326(5950), 289–93.
- Lin, K., Smit, S., Bonnema, G., Sanchez-Perez, G., & de Ridder, D. (2015). Making the difference: integrating structural variation detection tools. *Brief Bioinform*, 16(5), 852–64.
- Liu, D., Hunt, M., & Tsai, I. J. (2018). Inferring synteny between genome assemblies: a systematic evaluation. *BMC Bioinformatics*, 19(1), 26.
- Liu, G., & Zhang, T. (2021). Single Copy Oligonucleotide Fluorescence In Situ Hybridization Probe Design Platforms: Development, Application and Evaluation. *Int J Mol Sci*, 22(13), 7124.
- Liu, Z., Roesti, M., Marques, D., Hiltbrunner, M., Saladin, V., & Peichel, C. L. (2022). Chromosomal Fusions Facilitate Adaptation to Divergent Environments in Threespine Stickleback. *Mol Biol Evol*, 39(2).
- Logsdon, G. A., Vollger, M. R., Hsieh, P., Mao, Y., Liskovych, M. A., Koren, et al. (2021). The structure, function and evolution of a complete human chromosome 8. *Nature*, 593(7857), 101–7.

- Lucek, K., Gompert, Z., & Nosil, P. (2019).** The role of structural genomic variants in population differentiation and ecotype formation in *Timema cristinae* walking sticks. *Mol Ecol*, 28(6), 1224–37.
- Luo, M. C., Deal, K. R., Akhunov, E. D., Akhunova, A. R., Anderson, O. D., Anderson, J. A., et al. (2009).** Genome comparisons reveal a dominant mechanism of chromosome number reduction in grasses and accelerated genome evolution in Triticeae. *Proc Natl Acad Sci U S A*, 106(37), 15780–5.
- Lysák, M. A. (2022).** Celebrating Mendel, McClintock, and Darlington: On end-to-end chromosome fusions and nested chromosome fusions. *Plant Cell*, 34(7), 2475–91.
- Lysák, M. A., & Schubert, I. (2013).** Mechanisms of Chromosome Rearrangements. In *Plant Genome Diversity Volume 2* (pp. 137–47). Springer Vienna.
- Mahmoud, M., Gobet, N., Cruz-Dávalos, D. I., Mounier, N., Dessimoz, C., & Sedlazeck, F. J. (2019).** Structural variant calling: the long and the short of it. *Genome Biology*, 20(1), 246.
- Mandáková, T., & Lysák, M. A. (2018).** Post-polyploid diploidization and diversification through dysploid changes. *Curr Opin Plant Biol*, 42, 55–65.
- Mantere, T., Neveling, K., Pebrel-Richard, C., Benoist, M., van der Zande, G., Kater-Baats, E., et al. (2021).** Optical genome mapping enables constitutional chromosomal aberration detection. *Am J Hum Genet*, 108(8), 1409–22.
- Mardis, E. R. (2008).** The impact of next-generation sequencing technology on genetics. *Trends Genet*, 24(3), 133–41.
- Matthey, R. (1945).** L'évolution de la formule chromosomiale chez les Vertébrés. *Experientia*, 1(3), 78–86.
- King M. (1993):** *Species Evolution: The Role of Chromosome Change*. Cambridge University Press: Cambridge, UK 1993.
- Maxam, A. M., & Gilbert, W. (1977).** A new method for sequencing DNA. *Proc Natl Acad Sci U S A*, 74(2), 560–4.
- Mayrose, I., & Lysák, M. A. (2021).** The Evolution of Chromosome Numbers: Mechanistic Models and Experimental Approaches. *Genome Biol Evol*, 13(2).
- McCarthy, A. (2010).** Third Generation DNA Sequencing: Pacific Biosciences' Single Molecule Real Time Technology. *Chemistry & Biology*, 17(7), 675–6.
- McKinley, K. L., & Cheeseman, I. M. (2016).** The molecular basis for centromere identity and function. *Nat Rev Mol Cell Biol*, 17(1), 16–29.
- Mérot, C., Oomen, R. A., Tigano, A., & Wellenreuther, M. (2020).** A Roadmap for Understanding the Evolutionary Significance of Structural Genomic Variation. *Trends Ecol Evol*, 35(7), 561–72.
- Metzker, M. L. (2010).** Sequencing technologies — the next generation. *Nat Rev Genetics*, 11(1), 31–46.
- Meyer, A., Schloissnig, S., Franchini, P., Du, K., Woltering, J. M., Irisarri, I., et al. (2021).** Giant lungfish genome elucidates the conquest of land by vertebrates. *Nature*, 590(7845), 284–89.
- Miga, K. H., Koren, S., Rhie, A., Vollger, M. R., Gershman, A., Bzikadze, A., et al. (2020).** Telomere-to-telomere assembly of a complete human X chromosome. *Nature*, 585(7823), 79–84.
- Misteli, T. (2020).** The Self-Organizing Genome: Principles of Genome Architecture and Function. *Cell*, 183(1), 28–45.
- Morin, S. J., Eccles, J., Iturriaga, A., & Zimmerman, R. S. (2017).** Translocations, inversions and other chromosome rearrangements. *Fertil Steril*, 107(1), 19–26.
- Muñoz, A., & Sankoff, D. (2012).** Detection of gene expression changes at chromosomal rearrangement breakpoints in evolution. *BMC Bioinformatics*, 13(S3), S6.
- Nadeau, J. H., & Taylor, B. A. (1984).** Lengths of chromosomal segments conserved since divergence of man and mouse. *Proc Natl Acad Sci U S A*, 81(3), 814–8.
- Nagarajan, N., & Pop, M. (2013).** Sequence assembly demystified. *Nat Rev Genet*, 14(3), 157–67.
- Navarro, A., & Barton, N. H. (2003).** Chromosomal Speciation and Molecular Divergence-- Accelerated Evolution in Rearranged Chromosomes. *Science*, 300(5617), 321–4.

- Navarro, A., Betrán, E., Barbadilla, A., & Ruiz, A. (1997). Recombination and Gene Flux Caused by Gene Conversion and Crossing Over in Inversion Heterokaryotypes. *Genetics*, 146(2), 695–709.
- Neill NJ, Ballif BC, Lamb AN, Parikh S, Ravnan JB, et al, 2011: Recurrence, submicroscopic complexity, and potential clinical relevance of copy gains detected by array CGH that are shown to be unbalanced insertions by FISH. *Genome Res* 21:535-544.
- Ni, Y., Liu, X., Simeneh, Z. M., Yang, M., & Li, R. (2023). Benchmarking of Nanopore R10.4 and R9.4.1 flow cells in single-cell whole-genome amplification and whole-genome shotgun sequencing. *Comput Struct Biotechnol J*, 21, 2352–64.
- Norris, A. L., Workman, R. E., Fan, Y., Eshleman, J. R., & Timp, W. (2016). Nanopore sequencing detects structural variants in cancer. *Cancer Biol Ther* 17(3), 246–53.
- Nurk, S., Koren, S., Rhie, A., Rautiainen, M., Bzikadze, A. V., Mikheenko, A., Vollger, M. R., et al. (2022). The complete sequence of a human genome. *Science*, 376(6588), 44–53.
- O'Connor, R. E., Fonseka, G., Frodsham, R., Archibald, A. L., Lawrie, M., Walling, G. A., et al. (2017). Isolation of subtelomeric sequences of porcine chromosomes for translocation screening reveals errors in the pig genome assembly. *Anim Genet*, 48(4), 395–403.
- Pal, K., Forcato, M., & Ferrari, F. (2019). Hi-C analysis: from data generation to integration. *Biophys Rev*, 11(1), 67–78.
- Pardue, M. Lou, & Gall, J. G. (1969). Molecular Hybridization Of Radioactive Dna To The Dna Of Cytological Preparations. *Proceedings of the National Academy of Sciences*, 64(2), 600–4.
- Pellestor, F., Anahory, T., Lefort, G., Puechberty, J., Liehr, T., Hedon, B., et al. (2011). Complex chromosomal rearrangements: origin and meiotic behavior. *Hum Reprod Update*, 17(4), 476–94.
- Peona, V., Blom, M. P. K., Xu, L., Burri, R., Sullivan, S., Bunikis, I., et al. (2021). Identifying the causes and consequences of assembly gaps using a multiplatform genome assembly of a bird-of-paradise. *Mol Ecol Resour*, 21(1), 263–86.
- Peona, V., Weissensteiner, M. H., & Suh, A. (2018). How complete are “complete” genome assemblies?—An avian perspective. *Mol Ecol Resour*, 18(6), 1188–95.
- Perry, J., Slater, H. R., & Choo, K. H. A. (2004). Centric fission – simple and complex mechanisms. *Chromosome Res*, 12(6), 627–40.
- Pevzner, P., & Tesler, G. (2003). Transforming men into mice. *Proceedings of the Seventh Annual International Conference on Res Comput Mol Biol*, 247–56.
- Poot, M., & Hochstenbach, R. (2021). Prevalence and Phenotypic Impact of Robertsonian Translocations. *Mol Syndromol*, 12(1), 1–11.
- Rabanal, F. A., Gräff, M., Lanz, C., Fritschi, K., Llaca, V., Lang, M., et al. (2022). Pushing the limits of HiFi assemblies reveals centromere diversity between two *Arabidopsis thaliana* genomes. *Nucleic Acids Res*, 50(21), 12309–27.
- Ramos, L., & Antunes, A. (2022). Decoding sex: Elucidating sex determination and how high-quality genome assemblies are untangling the evolutionary dynamics of sex chromosomes. *Genomics*, 114(2), 110277.
- Rang, F. J., Kloosterman, W. P., & de Ridder, J. (2018). From squiggle to basepair: computational approaches for improving nanopore sequencing read accuracy. *Genome Biol*, 19(1), 90.
- Raudsepp, T., & Chowdhary, B. P. (2001). Correspondence of human chromosomes 9, 12, 15, 16, 19 and 20 with donkey chromosomes refines homology between horse and donkey karyotypes. *Chromosome Res*, 9(8), 623–9.
- Raudsepp, T., & Chowdhary, B. P. (2016). Chromosome Aberrations and Fertility Disorders in Domestic Animals. *Annu Rev Anim Biosci* 4(1), 15–43.
- Ravinet, M., Faria, R., Butlin, R. K., Galindo, J., Bierne, N., Rafajlović, M., et al. (2017). Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. *J Evol Biol*, 30(8), 1450–77.
- Rhie, A., McCarthy, S. A., Fedrigo, O., Damas, J., Formenti, G., Koren, S., et al. (2021). Towards complete and error-free genome assemblies of all vertebrate species. *Nature*, 592(7856), 737–46.

- Rhoads, A., & Au, K. F. (2015).** PacBio Sequencing and Its Applications. *Genomics Proteomics Bioinformatics*, 13(5), 278–89.
- Ritz, A., Bashir, A., Sindi, S., Hsu, D., Hajirasouliha, I., & Raphael, B. J. (2014).** Characterization of structural variants with single molecule and hybrid sequencing approaches. *Bioinformatics*, 30(24), 3458–66.
- Robertson, W. R. B. (1916).** Chromosome studies. I. Taxonomic relationships shown in the chromosomes of tettigidae and acrididae: V-shaped chromosomes and their significance in acrididae, locustidae, and gryllidae: Chromosomes and variation. *J Morphol*, 27(2), 179–331.
- Sacerdot, C., Louis, A., Bon, C., Berthelot, C., & Roest Crollius, H. (2018).** Chromosome evolution at the origin of the ancestral vertebrate genome. *Genome Biol*, 19(1), 166.
- Sahajpal, N. S., Barseghyan, H., Kolhe, R., Hastie, A., & Chaubey, A. (2021).** Optical Genome Mapping as a Next-Generation Cytogenomic Tool for Detection of Structural and Copy Number Variations for Prenatal Genomic Analyses. *Genes*, 12(3), 398.
- Sanger, F., & Coulson, A. R. (1975).** A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J Mol Biol*, 94(3), 441–8.
- Sankoff, D. (2009).** The where and wherefore of evolutionary breakpoints. *J Biol*, 8(7), 66.
- Sato, H., Masuda, F., Takayama, Y., Takahashi, K., & Saitoh, S. (2012).** Epigenetic Inactivation and Subsequent Heterochromatinization of a Centromere Stabilize Dicentric Chromosomes. *Current Biology*, 22(8), 658–67.
- Schadt, E. E., Turner, S., & Kasarskis, A. (2010).** A window into third-generation sequencing. *Hum Mol Genet*, 19(R2), R227–40.
- Schloss, J. A. (2008).** How to get genomes at one ten-thousandth the cost. *Nat Biotechnol*, 26(10), 1113–15.
- Schmid, M., Steinlein, C., Bogart, J. P., Feichtinger, W., León, P., La Marca, E., et al. (2010).** The Chromosomes of Terraranan Frogs. Insights into Vertebrate Cytogenetics. *Cytogenet Genome Res*, 130–131(1–8), 1–14.
- Schubert, I. (2018).** What is behind “centromere repositioning”? *Chromosoma*, 127(2), 229–234.
- Schubert, I., & Lysák, M. A. (2011).** Interpretation of karyotype evolution should consider chromosome structural constraints. *Trends Genet*, 27(6), 207–16.
- Schwartz, D. C., Li, X., Hernandez, L. I., Ramnarain, S. P., Huff, E. J., & Wang, Y.-K. (1993).** Ordered Restriction Maps of *Saccharomyces cerevisiae* Chromosomes Constructed by Optical Mapping. *Science*, 262(5130), 110–14.
- Sedlazeck, F. J., Lee, H., Darby, C. A., & Schatz, M. C. (2018)a.** Piercing the dark matter: bioinformatics of long-range sequencing and mapping. *Nat Rev Genet*, 19(6), 329–46.
- Sedlazeck, F. J., Rescheneder, P., Smolka, M., Fang, H., Nattestad, M., von Haeseler, A., et al (2018)b.** Accurate detection of complex structural variations using single-molecule sequencing. *Nat Methods*, 15(6), 461–8.
- Shizuya, H., & Kouros-Mehr, H. (2001).** The development and applications of the bacterial artificial chromosome cloning system. *Keio J Med* 50(1), 26–30.
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015).** BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31(19), 3210–2.
- Slatko, B. E., Gardner, A. F., & Ausubel, F. M. (2018).** Overview of Next-Generation Sequencing Technologies. *Curr Protoc Mol Biol*, 122(1).
- Slijepcevic, P. (1998).** Telomeres and mechanisms of Robertsonian fusion. *Chromosoma*, 107(2), 136–40.
- Srinivasan, S., & Batra, J. (2014).** Four Generations of Sequencing- Is it Ready for the Clinic Yet? *Next Gener Seq Appl*, 01(01).
- Sturtevant, A. H. (1917).** Genetic Factors Affecting the Strength of Linkage in *Drosophila*. *Proc Natl Acad Sci U S A*, 3(9), 555–8.

- Sun, L., Gao, T., Wang, F., Qin, Z., Yan, L., Tao, W., et al. (2020).** Chromosome-level genome assembly of a cyprinid fish *Onychostoma macrolepis* by integration of nanopore sequencing, Bionano and Hi-C technology. *Molecular Ecology Resources*, 20(5), 1361–71.
- Teague, B., Waterman, M. S., Goldstein, S., Potamouisis, K., Zhou, S., Reslewic, S., et al. (2010).** High-resolution human genome structure by single-molecule analysis. *Proc Natl Acad Sci U S A*, 107(24), 10848–53.
- Tian, H.-F., Hu, Q., Lu, H.-Y., & Li, Z. (2022).** Chromosome-Scale, Haplotype-Resolved Genome Assembly of Non-Sex-Reversal Females of Swamp Eel Using High-Fidelity Long Reads and Hi-C Data. *Front Genet* 13.
- Tian, H.-F., Hu, Q.-M., & Li, Z. (2021).** A high-quality *de novo* genome assembly of one swamp eel (*Monopterus albus*) strain with PacBio and Hi-C sequencing data. *G3 Genes|Genomes|Genetics*, 11(1).
- Towbin, B. D., Gonzalez-Sandoval, A., & Gasser, S. M. (2013).** Mechanisms of heterochromatin subnuclear localization. *Trends Biochem Sci* 38(7), 356–63.
- Trask, B. J. (2002).** Human cytogenetics: 46 chromosomes, 46 years and counting. *Nat Rev Genet*, 3(10), 769–78.
- Treangen, T. J., & Salzberg, S. L. (2012).** Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat Rev Genetics*, 13(1), 36–46.
- Tsai, A. G., & Lieber, M. R. (2010).** Mechanisms of chromosomal rearrangement in the human genome. *BMC Genomics*, 11(Suppl 1), S1.
- Tuna, M., Amos, C. I., & Mills, G. B. (2019).** Molecular mechanisms and pathobiology of oncogenic fusion transcripts in epithelial tumors. *Oncotarget*, 10(21), 2095–111.
- van Dijk, E. L., Jaszczyszyn, Y., Naquin, D., & Thermes, C. (2018).** The Third Revolution in Sequencing Technology. *Trends Genet*, 34(9), 666–81.
- Vara, C., Paytuví-Gallart, A., Cuartero, Y., Álvarez-González, L., Marín-Gual, L., Garcia, F., et al. (2021).** The impact of chromosomal fusions on 3D genome folding and recombination in the germ line. *Nat Commun*, 12(1), 2981.
- Vollger, M. R., Dishuck, P. C., Sorensen, M., Welch, A. E., Dang, V., Dougherty, M. L., et al. (2019).** Long-read sequence and assembly of segmental duplications. *Nature Methods*, 16(1), 88–94.
- Vollger, M. R., Logsdon, G. A., Audano, P. A., Sulovari, A., Porubsky, D., Peluso, P., et al. (2020).** Improved assembly and variant detection of a haploid human genome using single-molecule, high-fidelity long reads. *Ann Hum Genet*, 84(2), 125–40.
- Weckselblatt, B., & Rudd, M. K. (2015).** Human Structural Variation: Mechanisms of Chromosome Rearrangements. *Trends Genet*, 31(10), 587–99.
- Wellband, K., Mérot, C., Linnansaari, T., Elliott, J. A. K., Curry, R. A., & Bernatchez, L. (2019).** Chromosomal fusion and life history-associated genomic variation contribute to within-river local adaptation of Atlantic salmon. *Mol Ecol*, 28(6), 1439–59.
- Wellenreuther, M., & Bernatchez, L. (2018).** Eco-Evolutionary Genomics of Chromosomal Inversions. *Trends Ecol Evol*, 33(6), 427–40.
- Wenger, A. M., Peluso, P., Rowell, W. J., Chang, P.-C., Hall, R. J., Concepcion, G. T., Ebler, J., et al. (2019).** Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nature Biotechnology*, 37(10), 1155–62.
- Whiteford, N., Skelly, T., Curtis, C., Ritchie, M. E., Löhr, A., Zaranek, A. W., et al. (2009).** Swift: primary data analysis for the Illumina Solexa sequencing platform. *Bioinformatics*, 25(17), 2194–99.
- Wilhelm, K., Pentzold, C., Schoener, S., Arakelyan, A., Hakobyan, A., Mrasek, K., & Weise, A. (2018).** Fragile Sites as Drivers of Gene and Genome Evolution. *Curr Genet Med Rep*, 6(4), 136–43.
- Wingett, S. W., Ewels, P., Furlan-Magaril, M., Nagano, T., Schoenfelder, S., Fraser, P., et al. (2015).** HiCUP: pipeline for mapping and processing Hi-C data. *FI000Resh*, 4, 1310.

- Xia, L. C., Ai, D., Lee, H., Andor, N., Li, C., Zhang, N. R., et al. (2018).** SVEngine: an efficient and versatile simulator of genome structural variations with features of cancer clonal evolution. *GigaScience*, 7(7).
- Yang, L. (2020).** A Practical Guide for Structural Variation Detection in the Human Genome. *Curr Protoc Hum Gene*, 107(1).
- Yekefenhazi, D., He, Q., Wang, X., Han, W., Song, C., & Li, W. (2022).** Chromosome-level genome assembly of Nibea coibor using PacBio HiFi reads and Hi-C technologies. *Scientific Data*, 9(1), 670.
- Yin, Y., Fan, H., Zhou, B., Hu, Y., Fan, G., Wang, J., et al. (2021).** Molecular mechanisms and topological consequences of drastic chromosomal rearrangements of muntjac deer. *Nat Commun*, 12(1), 6858.
- Yoshida, K., Rödelsperger, C., Röseler, W., Riebesell, M., Sun, S., Kikuchi, T., et al. (2023).** Chromosome fusions repatterned recombination rate and facilitated reproductive isolation during *Pristionchus* nematode speciation. *Nat Ecol Evol*.
- Yu, S., & Graf, W. D. (2010).** Telomere Capture as a Frequent Mechanism for Stabilization of the Terminal Chromosomal Deletion Associated with Inverted Duplication. *Cytogenet Genome Res*, 129(4), 265–74.
- Yuan, Y., Bayer, P. E., Batley, J., & Edwards, D. (2021).** Current status of structural variation studies in plants. *Plant Biotechnol J*, 19(11), 2153–63.
- Yuan, Y., Chung, C. Y.-L., & Chan, T.-F. (2020).** Advances in optical mapping for genomic research. *Comput Struct Biotechnol J*, 18, 2051–62.
- Zhang, F., Khajavi, M., Connolly, A. M., Towne, C. F., Batish, S. D., & Lupski, J. R. (2009).** The DNA replication FoSTeS/MMBIR mechanism can generate genomic, genic and exonic complex rearrangements in humans. *Nat Genet*, 41(7), 849–53.
- Zhang, W., Friebe, B., Gill, B. S., & Jiang, J. (2010).** Centromere inactivation and epigenetic modifications of a plant chromosome with three functional centromeres. *Chromosoma*, 119(5), 553–63.
- Zhou, Y., Yang, L., Han, X., Han, J., Hu, Y., Li, F., et al. (2022).** Assembly of a pangenome for global cattle reveals missing sequences and novel structural variations, providing new insights into their diversity and evolutionary history. *Genome Research*, 32(8), 1585–601.
- Zeitouni, B., Boeva, V., Janoueix-Lerosey, I., Loeillet, S., Legoix-né, P., Nicolas, A., et al. (2010).** SVDetect: a tool to identify genomic structural variations from paired-end and mate-pair sequencing data. *Bioinformatics*, 26(15), 1895–6.

## Internetové zdroje

1. <https://www.bmkgene.com/dnarna-sequencing-illumina-sequencer-product/>
2. <https://www.pacb.com/technology/hifi-sequencing/>
3. <https://bionanogenomics.com/wp-content/uploads/2018/04/30206-Bionano-Prep-Direct-Label-and-Stain-DLS-Protocol.pdf>
4. <https://vertebrategenomesproject.org/phase-one>