

Tämän teoksen tekijänoikeudet kuuluvat CSC - Tieteellinen laskenta Oy:lle. Teoksen tai osia siitä voi kopioida ja tulostaa vapaasti henkilökohtaiseen käyttöön sekä Suomen yliopistojen ja korkeakoulujen käyttöön edellyttäen, että kopioon tai tulosteeseen liitetään tämä ilmoitus teoksen tekijänoikeuksista. Teosta ei saa myydä tai sisällyttää osaksi muita teoksia ilman CSC:n lupaa.

1 Johdanto

Bio- ja tietoteknologian kehittyminen avaa jatkuvasti uusia mahdollisuuksia biotieteiden tutkimuksessa. Biologisen ja lääketieteellisen informaation tietokoneavusteisesta keräämisestä, prosessoinnista ja analysoinnista on syntynyt uusi kasvava tieteenala, jota kutsutaan bioinformatiikaksi. Bioinformatiikan kirjallisuutta tai muuta opetusmateriaalia on julkaistu vasta vähän ja terminologia on epäyhdenäistä. Tästä syystä suomalaisista bioinformatiikka-asiantuntijoista koostuva Bioverkko päätti koota bioinformatiikan keskeistä sanastoa tukemaan bioinformatiikan opetusta sekä uusia bioinformatiikan käyttäjiä ja tutkijoita. Tavoitteena oli, että sanasto toimisi osaltaan siltana bioinformatiikassa yhdistyvien tieteenalojen välillä.

Bioinformatiikan sanaston ensimmäiseen versioon on pyritty kokoamaan tärkeimmät bioinformatiikan termit. Lisäksi mukaan on otettu bioinformatiikkaan läheisesti liittyvien tieteenalojen ja tekniikoiden (molekyylibiologia, tietojenkäsittely) termejä, jotka esiintyvät usein bioinformatiikan yhteydessä. Sanastoon on myös koottu joukko yleisesti käytettyjä bioinformatiikan algoritmeja, ohjelmistoja ja tietokantoja. Näiden suhteen sanastoa voidaan pitää vain suuntaa-antavana. Ohjelmistoja, tietokantoja ja verkkopalveluita on tarjolla niin runsaasti, että objektiivinen tärkeimpien bioinformatiikan työkalujen ja tietolähteiden valikoiminen näin lyhyeen listaan lienee mahdotonta.

Haluamme kiittää kaikkia sanaston valmistelussa avustaneita henkilöitä. Erityisesti Manne Miettistä sanaston XML-sivujen tuottamisesta. Jouni Aspia, Juha Haatajaa, Liisa Laakkosta, Tapio Salakoskea, Pekka Uimaria, Esko Ukkosta ja Imre Västrikiä haluamme kiittää sanastosta annetusta palautteesta.

Kaikki sanastoa koskeva palaute on tervetullutta. Sanastoa kehitetään edelleen ja se pyritään päivittämään saadun palautteen perusteella noin kerran vuodessa. Lisättäessä sanastoon uusia termejä tulee pääpaino olemaan bioinformatiikan termeissä. Ohjelmistojen ja palvelinten nimistöä, samoin kuin tietotekniikan ja molekyylibiologian termejä lisätään vain, mikäli ne katsotaan bioinformatiikan kannalta erityisen tärkeiksi. Palautteen voi lähettää Kimmo Mattilalle osoitteeseen Kimmo.Mattila@csc.fi.

Espoossa 5.12. 2000
Sanastotyöryhmä

2 Kaikki termit aakosjärjestyksessä

Numerot ja symbolit

(-)-strand (-)-juoste (DNA:ssa), ((-)-nauha, (-)-säie), templaattijuoste Syntyvä juoste on templaatille komplementaarinen.

(+)-strand (+)-juoste (DNA:ssa) ((+)-nauha, (+)-säie) Koodaava juoste, jonka kaltainen syntyvä juoste tulee olemaan. Vastinjuoste, joka ei ole templaattijuoste. (+)-juosteella ja lähettiRNA:lla on sama sekvenssi (poikkeuksena T:n korvaus U:lla).

3'-end 3'-pää nukleinihappojuosteessa Merkintä 3' viittaa sokeriryhmän hiiliatomien numerointiin. Katso kuvaa 2

3'-splice site 3'-silmutointikohta

5'-end 5'-pää nukleinihappojuosteessa Merkintä 5' viittaa sokeriryhmän hiiliatomien numerointiin. Katso kuvaa 2

5'-splice site 5'-silmutointikohta

A

Ab initio *lat.* Alusta, perusteista Laskennassa ab initio menetelmillä pyritään löytämään ratkaisu annetun teorian perusteella, ilman kokeellisen aineiston ohjausta. Esimerkiksi proteiinin 3-ulotteisen rakenteen ennustaminen puhtaasti teoreettisen mallin avulla.

acceptor *akseptori, vastaanottaja* Atomi, molekyyli tai solu, joka vastaanottaa luovuttajalta esimerkiksi elektronin, molekyylin tai molekyyliyhdyntien

acceptor-(3')-splice site *akseptori-3'-silmutointikohta*

A-DNA A-DNA B-DNA:sta dehydraation kautta syntynyt rakennemuoto, jossa on 11 emäsparia/kierre. Katso B-DNA ja Z-DNA.

alignment *rinnastus (kohdistus, linjaus)* Sekvenssien vastinmerkkien tai vastinkohtien sijoittaminen kohdakkain.

allele *alleeli, vastingeeni* Geenin vaihtoehtoiset muodot, jotka sijaitsevat samassa kohdassa kromosomissa. Alleeleja voi samalla diploidilla normaalikaryotyypillä yksilöllä olla kerrallaan vain kaksi,

joko kaksi samanlaista (homotsygootti) tai kaksi erilaista (heterotsygootti).

α -carbon *α -hiili* Katso selitystä aminohappojen osien nimeämisestä. Katso kuvaa 4

α -helix *α -kierre* Proteiinin sekundääri-rakenne, jossa proteiinipääketju muodostaa kierteen. Kierrettä pitävät koossa ensimmäisen (n) peptidisidoksen karbonyyliryhmän (C=O) ja neljännen (n+4) peptidisidoksen amidiryhmän (H-N) väliset vetysidokset.

alternative splicing *vaihtoehtoinen silmutointi, pujonta* Proteiinituotannon säätely lähettiRNA:n prosessoinnilla. Eksonien koodaamien lähettiRNA-sekvenssien liittäminen toisiinsa vaihtoehtoisista kohdista saa aikaan erilaisten proteiinien tuotannon, tai lähettiRNA:n, jota ei transloida.

Alu sequence *Alu-sekvenssi, yleinen toistojakso* Useissa ihmisen introneissa toistuva sekvenssi, joka tunnustetaan AluI-restriktiokohtien avulla. Homologiselle rekombinaatiolle alttiita kohtia.

AMBER Biologisille makromolekyyille kehitetty mallitusohjelmisto ja voimakenttä. <http://www.amber.ucsf.edu/amber/amber.html>

amber mutation *amber-mutaatio* Lope-tuskodoniin (UAG) johtava mutaatio.

amphipathic *amfipaattinen* Molekyyli tai sen osa on amfipaattinen, mikäli sen pinnalla on selkeästi erillisiä vesihakuisia ja vesipakoisia alueita.

amphipathic helix *amfipaattinen kierre* Proteiinin rakenteessa esiintyvä kierrakerakenne, jossa vesihakuiset ja vesipakoiset aminohapot sijoittuvat kierteen vastakkaisille puolille.

amplification *vahvistus, suurennus; erityisesti PCR:n yhteydessä myös monistaminen*

annealing (hybridization) 1. Molekyyli-biologiassa: komplementaaristen yksin-

auhaisten nukleiinihappojen pariutumisen. 2. Numeriikassa: Simulated Annealing, laskennallinen menetelmä, jota käytetään mm. optimointitehtävissä.

annotation *sekvenssikuvaus, annotaatio*. Sekvenssitietokantojen yhteydessä: sekvenssin ominaisuuksien kuvaaminen varsinaisen sekvenssin lisäksi tietokannassa. Genomisekvenssistä voidaan määrittää tunnettuja ja ennustettuja geenejä, näiden koodaavia alueita ja säätelyalueita, variaatioita, STS-kohtia jne. Proteiinisekvenssin kuvaus voi puolestaan sisältää tietoja proteiinin toiminnasta, translaation jälkeisistä modifikaatioista, domeineista, aktiivisista kohdista ja rakenteesta.

anticodon *antikodoni* Siirtäjä-RNA:n (tRNA:n) kolmen peräkkäisen kodonille vastakkaisen eli komplementaarisen nukleotidin muodostama tripletti, joka koodaa yhtä aminohappoa. Antikodoni tunnistaa kodonin, mikä mahdollistaa aminohappojen sijoittamisen oikealle kohdalle peptidiketjussa.

anticodon arm *antikodonisilmukka* tRNA:n antikodonin sisältävä osa.

antiparalleel *vastakkaissuuntainen, antiparalleelinen*

antisense Geenille vastakkainen DNA-sekvenssi, joka toimii RNA:n mallina transkriptiossa.

autosome *autosomi* Muu kuin sukupuolikromosomi.

B

BAC (bacterial artificial chromosome) katso bacterial artificial chromosome.

backbone *päätetju (selkäranka, ranka)*

bacterial artificial chromosome (BAC) *keinotekoinen bakteerikromosomi* Molekyylibiologiassa käytetty kantaja (vektori).

bait *koetin* Katso: probe

base *emäs* Nukleiinihappojen yhteydessä: Nukleiinihappojen perusyksikkö. DNA:n neljä emästyyppeä ovat adeniini (A),

guaniini (G), tymiini (T) ja sytosiini (C). RNA:ssa tymiinin tilalla esiintyy urasiili (U).

base pair (bp) *emäspari* Nukleotidien emäkset, jotka kaksijuosteisessa DNA:ssa sitoutuvat toisiinsa vetysidoksilla. DNA:n neljä emästä ovat adeniini (A), guaniini (G), tymiini (T) ja sytosiini (C). Emäspareilla tarkoitetaan yleensä pareja A-T ja G-C, jotka ovat energeettisesti edullisimmat. RNA:ssa T korvautuu urasiililla (U).

Bayes' rule *Bayesin sääntö* Bayesilaisen todennäköisyyslaskennan perusolettamus. Bayesin sääntöä käytetään arvioitaessa ehdollisia todennäköisyyksiä, eli kahden tai useamman parametrin vaikutusta jonkin tapahtuman todennäköisyyteen. Esim. kahden parametrin (A ja B) tapauksessa sääntö on muotoa: Tapahtuman A ehdollinen todennäköisyys tapahtuman B yhteydessä, $P(A|B)$ on yhtä suuri kuin tapahtuman A kokonaistodennäköisyys, $P(A)$, kertaa tapahtuman B ehdollinen todennäköisyys tapahtuman A yhteydessä, $P(B|A)$, jaettuna B:n kokonaistodennäköisyydellä, $P(B)$. $P(A|B) = P(A)P(B|A) / P(B)$

Bayesian *Bayesilainen* Menetelmä, jossa käytetään Bayesin sääntöä.

B-DNA *B-DNA* DNA:n hydroitu rakenemuoto (Vertaa: A-DNA ja Z-DNA). Watsonin ja Crickin kuvaama DNA-rakennemuoto, jossa on 10,4 emäsparia/kierre.

β -hairpin *hiusneularakenne* Proteiinin sekundäärirakenne, joka koostuu kahdesta vastakkaissuuntaisesta β -säikeestä ja niitä yhdistävästä silmukasta.

β -sheet *β -levy, β -taso* Vierekkäisten β -säikeiden muodostama vetysidosten stabiloima proteiinien sekundäärirakenne. Taso on yhdensuuntainen tai vastakkaissuuntainen sen mukaan, kulkevatko vierekkäisten säikeiden peptidiketjut saman- vai erisuuntaisesti.

β -strand *β -säie* Proteiinin sekundäärirakenne, jossa päätetju on lähes mak-

simaalisesti oienneena. β -säikeet voivat esiintyä yksittäin, mutta yleensä ne sijoittuvat proteiinirakenteissa vierekkäin muiden β -säikeiden kanssa muodostaen β -levyjä.

β -turn β -käännös Proteiinin sekundaärirakenne, jossa peptidiketju kääntyy jyrkästi kolmen aminohapon matkalla. Vetysidos pitää käännöstä koossa. β -käännöksiä tunnetaan useita eri tyyppinä.

bioinformatics bioinformatiikka Biologisen ja lääketieteellisen informaation tietokoneavusteinen kerääminen, prosessointi ja analysointi.

BLAST (Basic Local Alignment Search Tool) Ohjelma/algoritmi joka etsii hakusekvenssin kanssa samankaltaista sekvenssiä ja rinnastaa sekvenssit paikallisesti. Hakusekvenssi ja -tietokanta voivat olla nukleotidi- ja proteiinisekvenssien mikä tahansa yhdistelmä eli DNA-sekvenssillä voi hakea myös aminohaposekvenssitietokannasta ja päin vastoin. <http://www.ncbi.nlm.nih.gov/BLAST/>

BLITZ Rinnakkaislaskentaa hyödyntävä Smithin ja Watermanin algoritmiin pohjautuva sekvenssihakuohjelma.

BLOCKS Tietokanta ja analyysipalvelin hyvin samankaltaisten proteiinisekvenssi-alueiden etsintään ja vertailuun. <http://www.blocks.fhrc.org/>

BLOSUM (Blocks Substitution Matrix) Pisteytysmatriisisarja, jota käytetään sekvenssianalyseissa. Perustuu BLOCKS-tietokannan kaukaista sukua olevien sekvenssien paikallisiin rinnastuksiin.

blunt ended tylppäpäinen Kaksinauhaisen DNA:n pää, jossa kumpikin nauha on yhtä pitkä.

bootstrap itselataava Menetelmä, jossa algoritmin antamaa tulosta testataan toistamalla algoritmia sekoitetuilla aloitusarvoilla tai syöttötiedoilla. Sekoitus tehdään satunnaisesti, mutta kuitenkin tavalla, jonka ei pitäisi vaikuttaa käytetyn

menetelmän antamaan tulokseen.

bp (base pair) katso base pair.

byte (MB) *tavu* Määrämittainen, yleensä kahdeksan bitin muodostama kokonaisuus. Esimerkiksi yhden kirjainmerkin esittämiseen tarvitaan yleensä yksi tavu.

C

CAAT-box CAAT-sekvenssi Useiden eukaryoottigeenien promoottorialueella sijaitseva konservoitunut DNA-sekvenssi, joka yleensä osallistuu transkripti- on säätelyyn. CAAT-sekvenssi sijaitsee yleensä ns. TATA-laatikon 5'-puolella, n. 50 - 200 nukleotidia alavirtaan transkripti- on aloituskohdasta.

CAI (codon adaptation index) katso codon adaptation index.

candidate gene DNA-alue, jonka sijainti ja sekvenssi saattavat liittyä tarkasteltavaan ominaisuuteen, esim. perinnölliseen sairauteen.

cap Kemiallinen rakenne, joka liitetään aitotumallisten esi-lähtetri- NA:n 5'-päähän transkription jäl- keen. Modifioitu guanosiinitrifosfaatti- 7-metyyliguanosiini, joka on liittynyt 5'- 5' trifosfaattisidoksella lähtetriRNA:n 5'- päähän.

cap site lähtetriRNA:n 5'-pää. Kohta, johon cap-rakenne liitetään, merkitään usein geeniin +1:llä.

carrier kantaja 1. Taudinkantaja, yksilö, joka ei itse sairasta tautia, mutta voi välittää sen aiheuttajaa (bakteeria, virusta) muihin yksilöihin. 2. Heterotsygoottinen yksilö, jolla on kromosomistossaan peittyvä geeni (esim. tautia aiheuttava). 3. Kuljettava aine, kantaja-aine, vektori.

CATH (Class, Architecture, Topology and Homologous superfamily). Tietokanta, jossa PDB-rakenteiden proteiinidomeenit on luokiteltu rakenteellisten ominaisuuksien mukaan. <http://www.biochem.ucl.ac.uk/bsm/cath/>

cdNA (complementary DNA) *komplemen-*

taarinen DNA, vastakkainen DNA DNA-juoste, joka on syntetisoitu käyttäen mallina RNA:ta.

cDNA library *cDNA-kirjasto* Kokoelma cDNA-molekyylejä, jotka on pakattu kantajaan (plasmidi, faagi).

Celera Genomien sekvensointiin erikoistunut yhtiö. <http://www.celera.com/>

centimorgan (cM) *senttimorgan* Geneettisen etäisyyden mitta (1/100 Morgania). 1 cM vastaa karkeasti ottaen 1 miljoonaa emäsparia DNA:ta ja yhden prosentin rekombinaatiofraktiota. Rekombinaatiofraktion ja karttaetäisyyden tarkka suhde määritellään karttafunktioiden avulla.

centromere *sentromeeri* Kromosomikäsivarsien kiinnittymiskohta, johon tuman jakautumisessa kiinnittyvät myös tumasukkulan säikeet.

chaperone *kaperoni* Proteiini, joka avustaa toisen proteiinin laskostumista.

CHARMM Biologisille makromolekyyleille kehitetty mallitusohjelmisto ja voimakenttä. <http://yuri.harvard.edu/>

chiasma *kiasma* Meioosissa homologisten kromatidien välille muodostuva rakenne, joka mahdollistaa kromatidien geneettisen materiaalin vaihtuminen eli tekijäinvaihdunnan.

chip (oligonucleotide array) *geenilastu, DNA-siru* Lasi- tai piilevy, tai kalvo, jonka pinnalle on syntetisoitu tai sidottu DNA-molekyylejä (geenien osia tai oligonukleotideja). Käytetään esimerkiksi geenien ilmentymisen analysoimisessa. Katso DNA microarray/cDNA array.

Chou-Fasman analysis *Choun ja Fasmanin analyysi* Yksi proteiinien sekundaärirakenteen ennustusmenetelmistä.

chromatid *kromatidi* Kahdentuneen kromosomin puolikas eli kromosomin puolikkaat ennen solunjakautumista edeltävää mitoosin metafaasivaihetta tai meioosin toisen jaon loppuvaiheita.

chromatin *kromatiini* Kromosomin

pakkautunut DNA-rihma histoniproteiineineen. (Mikroskooppisesti: tiheästi pakkautunut heterokromatiini ja löysempi eukromatiini).

chromosomal map *kromosomikartta* Kromosomikuva, jossa on merkittynä geenimerkkien fyysinen sijainti.

chromosome *kromosomi* DNA-rihma, joka aitotumallisilla pakkautuu tiiviisti histoniproteiinien avulla. Pääosa solun geenistöstä on kromosomeissa.

clade *oksa, haara* sukupuun haara, kehityslinja. Samansukuinen ryhmä.

cladogram *kladogrammi* Puukaavio, joka kertoo tarkasteltavien yksiköiden eriytymisjärjestyksen, muttei erkaantumiseen kulunutta aikaa. Perustuvat kladistisiin ominaisuuksiin, joiden oletetaan sisältävän tietoa kehityshistoriallisesta alkuperästä.

clone *klooni* Perinnöllisesti samanlaisten yksilöiden tai solujen joukko.

Clustal (ClustalW, ClustalX) Monen sekvenssin (DNA- tai aminohappo-) rinnastusohjelma. http://www.infobiogen.fr/docs/ClustalW/clustalw_help.html

cM (centimorgan) katso centimorgan.

coalescent *yhteensulautuminen* Evoluutiobiologiassa: teoria, jossa nykyhetkestä palataan mutaatioiden kautta muinaisiin alleleihin. Voidaan käyttää esimerkiksi mutaatiotaajuuksien arvioinnissa.

coding strand koodaava juoste. Genomisen DNA:n juoste, jonka sekvenssi kopioituu lähettiRNA:han. Sekvenssiä julkaistaessa esitettävä emäsjärjestys.

codominance *kodominanssi* Periytymistapa, jossa molemmat alleelit ilmenevät fenotyypissä.

codon *kodoni* lähettiRNA:ssa kolme peräkkäistä emästä, jotka proteiinisynthesissä vastaavat yhtä aminohappoa tai translaation lopetusta.

codon adaptation index (CAI) *kodonien adaptaatioindeksi* Synonyymikodonien käyttöön liittyvä mitta.

cohesive end *koheesiivinen pää* Kaksijuosteisen DNA:n pää, jossa toinen juoste on pidempi eli jatkuu yksijuosteisena.

coiled coil *kiertynyt kierre* Proteiini-rakenne, jossa samansuuntaiset α -kierteet kiertyvät toistensa ympäri.

complementary *komplementaarinen* Vastintai nukleinihappojuoste, jonka jokainen nukleotidimäs pariutuu (Watson-Crick -periaatteen mukaisesti) toisen juosteen kanssa.

complementary base *vastinemäs* Kaksijuosteisessa DNA:ssa jokaista A:ta vastaa T ja jokaista G:tä C.

complex disease *monitekijäinen tauti* Tila, jonka syntyyn vaikuttavat useiden geenien lisäksi myös ympäristötekijät. Katso multifactorial ja multigenic disease sekä polygenic disorders.

consensus sequence *konsensussekvenssi* DNA-, RNA- tai aminohapposekvenssiryhmän perusteella matemaattisin menetelmin tehty sekvenssi, joka kuvastaa nukleotidien tai aminohappojen yleisintä muotoa kussakin sekvenssin kohdassa. Alueet, joilla vastaavuus on suuri, heijastavat yleensä yhteistä toiminnallista historiaa.

contig *jatkumo* Lyhyistä, osittain päällekkäisistä sekvensseistä yhdistelemällä muodostettu yhtenäinen sekvenssi.

cosmid *kosmidi* Kantaja (vektori), joka on tehty yhdistämällä λ -faagin DNA:ta bakteerin plasmidiin. Käytetään usein pitkien DNA-fragmenttien kloonauksessa.

covariance *kovarianssi* Kahden muuttujan yhteistä vaihtelua kuvaava tunnusluku.

CpG island *CpG-saareke* Runsaasti C- ja G-nukleotideja sisältävä alue. Usein lähellä aktiivisesti ilmennettäviä geenejä.

crossing over *tekijäinvaihdunta* Geneettisen materiaalin vaihtuminen vastinkromosomien välillä kromatidirihojen ristiinmenon yhteydessä meiosisin 1. jaossa.

CSC - Tieteellinen laskenta Oy Opetusministeriön omistama voittoa tavoit-

telematon osakeyhtiö, joka tarjoaa supertietokone-, laskenta-, ohjelmisto-, tietokanta- ja asiantuntijapalveluita mm. bioinformatiikassa. <http://www.csc.fi>

D

Dali Menetelmä ja WWW-palvelin, jolla voi verrata proteiinin 3D-rakennetta PDB-tietokannassa oleviin rakenteisiin. Katso myös FSSP. <http://www.ebi.ac.uk/dali/>

DD (differential display) katso differential display.

DDBJ (DNA DataBank of Japan) Japanilainen nukleotidisekvenssitietokanta, jota ristiinpäivitetään EMBL:n ja GenBankin kanssa. <http://www.ddbj.nig.ac.jp/Welcome.html>

deamination *deaminaatio* Aminoryhmän poisto.

degeneracy *degeneroituneisuus* Molekyylibiologiassa: Useimpia aminohappoja voi koodata useampi kuin yksi kodoni eli samaa aminohappoa voi vastata useampi kuin yksi nukleotidisekvenssi. Aminohapposekvenssistä ei välttämättä selviä alkuperäinen nukleotidisekvenssi, mutta nukleotidisekvenssi voidaan yksiselitteisesti kääntää aminohapposekvenssiksi. Tästä johtuen geneettinen koodi voidaan luokitella degeneroituneeksi koodiksi.

deletion *poistuma, deletio, puuttuma* Nukleinihappojuosteen osan ja siinä olevien geenien poistuminen. Mutaatityyppi.

denaturation *denaturoituminen* Nukleinihappojen vastinjuosteiden irtautuminen emäspariutumisen purkautuessa, tai proteiinien kolmiulotteisen rakenteen avautuminen ulkoisten tekijöiden vaikutuksesta (lämpötila, pH, jne.).

dendrogram *dendrogrammi, puukaavio* Mikä tahansa puumainen kaavio, joka kuvaa yksikköjen - kuten organismien - välisiä suhteita. Esim. sukupuu.

de novo *uudelleen, alusta* Synteesi yksinkertaisista rakennekomponenteista lähtien.

depurination *depurinaatio* Puriinin poisto nukleiinihapoista.

DGGE technique DNA-geelielektroforeesi denaturoivassa pH-gradientissa. Käytetään geenien monimuotoisuuden analysoinnissa.

dideoxynucleotide *dideoksinukleotidi* Nukleotidi, jossa deoksiriboosi on korvattu dideoksiriboosilla. DNA-polymeraasit eivät pysty liittämään uusia nukleotideja dideoksinukleotideihin, mistä johtuen niitä voidaan käyttää polymerisaation pysäyttämiseen.

dideoxynucleotide sequencing *dideoksinukleotidisekvensointi* Sangerin entsyymaattinen DNA:n sekvensointimenetelmä, jossa sekvensoitavalle juosteelle syntetisoidaan dideoksinukleotidien läsnäollessa leimattuja, eri pituisia juosteita. Pituusjärjestyksen avulla voidaan määrittää DNA:n emäsjärjestys.

differential display (DD) lähettiRNA-populaatioista RT-PCR:llä muodostettu kokoelma eri pituisia DNA-fragmentteja, jotka sekvensointigeelissä eroteltuna osoittavat samankaltaisesti tai eri tavoin ilmeneviä genejä.

dihedral angle *kiertokulma, tasokulma, dihedraalikulma*

diploid *diploidi* Solu, jonka tumassa on kaksi vastinkromosomistoa (yksi kummaltakin vanhemmalta), yleensä somaattiset eli muut kuin sukusolut.

diplotype *diplotyyppi* Kutakin kromosomia kaksi kappaletta.

direct repeat *suora toisto* Samassa suunnassa toistuva emäsjärjestys (esim. ACT ACT ACT). Katso palindromi.

Discover Kaupallinen, MSI:n tuottama molekyylibiologia-ohjelmisto, joka soveltuu myös biologisten makromolekyylien simuloimiseen.

distance geometry *etäisyysgeometria* Molekyylien rakennemäärittelyssä käytettävä laskennallinen menetelmä. Menetelmässä määritetään joukko atom-

ien välisten etäisyyksien välisiä raja-arvoja, joiden perusteella luodaan molekyylin rakennemalleja. Käytetään etenkin NMR-spektroskopian yhteydessä.

distance matrix *etäisyysmatriisi* Taulukko, joka kuvaa taksonien parittaisia eroja.

disulphide bond *disulfididisidos, rikkisilta* Kahden rikkiatomin välinen kovalenttinen sidos. Proteiineissa kysteiniinivuketjujen välisillä disulfididisidoksilla eli rikkisilloilla on usein merkittävä vaikutus proteiinin kolmiulotteiseen rakenteeseen. Katso disulphide bridge.

disulphide bridge *rikkisilta* Kahden kysteiniinivuketjun rikkiatomien välinen disulfididisidos. Proteiininrakenteissa rikkisillat voivat sitoa kovalenttisesti yhteen kaksi sekvenssin eri kohdissa olevaa kysteiniä ja siten vaikuttaa proteiinin rakenteeseen ja sen vakauteen.

DNA (deoxyribonucleic acid) *deoksiribonukleiinihappo*

DNA microarray *DNA-siru* Alustalle (lasi tai kalvo) kiinnitetty kokoelma DNA-molekyylejä, jotka voidaan hybridisoida liuoksessa olevan nukleiinihappokokoelman kanssa.

docking *telakointi* Molekyylien sovittaminen toisiinsa esimerkiksi ligandien sitoutumista tutkittaessa.

domain *domeeni* Proteiininrakenteen itsenäisesti poimuttava yksikkö.

donor *luovuttaja, donori* Atomi, molekyyli tai solu, joka luovuttaa vastaanottajalle esimerkiksi elektronin, molekyylin tai molekyyliyryhmän

double helix (Watson-Crick model)

kaksoiskierre, Watsonin ja Crickin malli Molekyylibiologiassa: DNA:n vastinjuosteet kiinnittyvät emäsosistaan toisiinsa kaksoiskierteiseksi molekyyliarakenteeksi.

double-stranded *kaksijuosteinen* (-nauhainen, -säikeinen) Nukleiinihapporakenne, joka muodostuu kahdesta emäspariutuneesta vastinjuosteesta.

Double Twist Bioinformatiikkapalveluja

tarjoava yritys. <http://www.doublet.wist.com/>

downstream *alavirran puoleinen, 5'-3' -suuntaan etenevä, 3'-puolella sijaitseva; proteiineissa karboksiterminaalipään puoleinen.*

dsDNA (double-stranded DNA) *kaksijusteinen (-nauhainen, -säikeinen) DNA*

dsRNA (double-stranded RNA) *kaksijusteinen (-nauhainen, -säikeinen) RNA*

duplication (*kromosomin osan kahdentuminen*) Kromosomin osan kahdentuminen. Mutaatiotyyppejä.

E

EBI (European Bioinformatics Institute) Euroopan bioinformatiikan instituutti Hinxtonissa, Cambridgessä (UK). <http://www.ebi.ac.uk>

edit distance *muokkausetäisyys, editointietäisyys*

EM algorithm (expectation maximisation algorithm) *odotuksen maksimointialgoritmi* Yleinen menetelmä maksimaalisen samankaltaisuuden laskemiseen.

EMBL Euroopan molekyylibiologian laboratorion ylläpitämä DNA-sekvenssitietokanta, jota ristiinpäivitetään DDBJ:n ja GenBankin kanssa. <http://www.ebi.ac.uk/embl/index.html>

EMBnet Tiedepohjainen eurooppalaisten (ja useiden Euroopan ulkopuolisten) asiantuntijakeskusten muodostama verkosto, jonka avulla kunkin osanottajamaan molekyylibiologit saavat palveluja kansallisen keskuksen tarjontaa laajemmasta alalta. <http://www.embnet.org>

EMBOSS (The European Molecular Biology Open Software Suite) Sekvenssianalytiikan ohjelmistopaketti. Eurooppalaisten bioohjelmistokehittäjien yhteistyöhanke, jossa yhdistetään jo olemassa olevia ja kehitettäviä ohjelmia ja työkaluja julkisesti saatavaksi ohjelma- ja kirjastokokonaisuudeksi. <http://www.uk.embnet.org/Software/EMBOSS>

endonuclease *endonukleaasi* Nukleiinihappoa juosteen keskeltä pilkkova entsyymi.

enhancer *tehostaja, voimistaja* Molekyylibiologiassa: transkriptiotekijä, joka yleensä aktivoi transkriptiota.

Ensembl EBI:n ja Sanger Centerin ylläpitämä genomidataan keskittynyt WWW-palvelin. <http://www.ensembl.org/>

EST (expressed sequence tag) katso expressed sequence tag.

eukaryote (eucaryote) *eukaryootti, aiotumallinen* Eliö, jonka kromosomeilla on nukleosomirakenne ja jonka kromosomit ovat tumassa tumakotelon eristämistä solulimasta. Solulimassa on tehtäviltään erikoistuneita soluelimiä.

Eulerian path/cycle *Eulerin polku/polku* Graafiteorian mukaisessa esityksessä polku, joka kulkee graafin kaikkien särmien (edge) kautta. Mikäli polku kulkee niin että se lopuksi palaa lähtöpisteeseensä, kyseessä on Eulerin ketju. Katso Graph theory.

evolutionary distance *evoluutioetäisyys* Taksonien väliset erot esim. molekyyllisissä ovat tarkasteltavissa stokastisin menetelmin, joiden avulla kullekin taksoniparille saadaan evolutiivista etäisyyttä kuvaava tunnusluku.

evolutionary tree *evoluutiopuu* Puukaavio, jolla esitetään taksonien evoluutioetäisyyttä.

exon *eksoni* Aiotumallisen geenin proteiinia koodaava osa, joita voi olla geenissä useita. Eksoneita on myös aiotumallisten viruksissa. Eksonien välisiä osia kutsutaan introneiksi.

exonuclease *eksonukleaasi* Nukleiinihappojuostetta päistä hajottava entsyymi.

exonuclease footprinting DNase footprinting Menetelmä, jolla voidaan paikantaa DNA:han sitoutuvien proteiinien sitoutumiskohta.

ExpPASy (Expert Protein Analysis System) Proteomiikkapalvelin, jota ylläpitää

Swiss Institute of Bioinformatics (SIB).
<http://www.expasy.ch/>

expressed sequence tag (EST) *ilmenevän geenin osa* EST-sekvenssi on solun tietyssä tilassa ilmenevän geenin satunnaisesti ja usein vain osittain sekvensoitu cDNA-kloonii.

expression *ilmentyminen* DNA:n koodaaman proteiinin tuottaminen. Termiä käytetään usein (virheellisesti) synonyyminä transkriptiolle.

extinction *sukupuutto*

ex vivo *kehon ulkopuolella*

F

FASTA Sekvenssihakuohjelma, joka perustuu Pearsonin ja Lipmanin algoritmiin.
<http://fasta.bioch.virginia.edu/>

FISH (fluorescence in situ hybridization) katso fluorescence in situ hybridization.

Fitch-Margoliash method *Fitch-Margoliash -menetelmä* Evolutiikassa: puukaavioiden luomisessa käytettävä heuristinen menetelmä.

flanking sequence kohdesekvenssin viereiset alueet (5'-flank, 3'-flank).

fluorescence in situ hybridization (FISH) Kudos- tai soluvalmisteille käytettävä hybridisaatioanalyysimenetelmä, jossa käytetään fluoresoivia merkkiaineita eli leimoja.

flush ended *kohesiivoinen pää* Kaksijuosteinen DNA, jossa toinen juosteista on pidempi. Katso sticky ended.

fold *yleisrakenne, poimutus* Käytetään kuvaamaan proteiinin tai muun makromolekyylin järjestäytyntä kolmiulotteista rakennetta yleisellä tasolla. Proteiineilla voi olla samanlainen yleisrakenne (esimerkiksi neljän kierteen kimppu), vaikka niiden aminohappokoostumus ja tarkka kolmiulotteinen rakenne ovat erilaisia.

folding *laskostuminen, poimuttuminen* Prosessi, jossa proteiinin tai muun makro-

molekyylin kolmiulotteinen rakenne hakeutuu sille ominaiseen muotoon.

frameshift Lukukehyksen muuttuminen esim. poistuman tai lisäyksen seurauksena.

FSSP (Fold classification based on Structure-Structure alignment of Proteins) Tietokanta, jossa PDB-rakenteet on luokiteltu 3D-rakenteidensa perusteella Dalimenetelmällä. <http://www.ebi.ac.uk/dali/fssp/fssp.html>

FTP (File Transport Protocol) Yleinen tiedostojen siirtokäytäntö tietokoneiden välillä.

G

gamete *sukupuolisolu* Tavallisesti haploidit solut, jotka yhtyessään muodostavat tsygootin.

gap *aukko* Aukko sekvenssirinnastuksessa.

GAP Ohjelma sekvenssien parittaiseen vertailuun.

GCG (The Wisconsin Package of the Genetics Computer Group, Inc.) Sekvenssianalytiikkaohjelmisto, joka sisältää ohjelmia DNA-, RNA- ja protei-nisekvenssien karakterisointiin, hakuun ja vertailuun. http://www.accelrys.com/products/gcg_wisconsin_package/index.html

GenBank Yhdysvaltalainen NCBI:n ylläpitämä DNA-sekvenssietietokanta, jota ristiinpäivitetään EMBL:n ja DDBJ:n kanssa. <http://www.ncbi.nlm.nih.gov/Genbank/index.html>

gene expression *geenin ilmeneminen* Puhkielessä (virheellisesti) usein transkriptio.

Genehunter Kytöntäanalyysiohjelma, jota voi käyttää sekä parametriseen että eiparametriseen kyöntäanalyysiin. Ohjelmalla voi analysoida vain pieniä sukupuita, mutta useita geenimerkkejä voidaan käyttää samanaikaisesti. <http://linkage.rockefeller.edu/soft/gh/>

genetic algorithm *geneettinen algoritmi*

Evoluutiomekanismeja matkiva optimointimenetelmä. Algoritmi perustuu ongelman kuvaamiseen ratkaisujoukkona, johon kohdistetaan valintaa, tekijäinvaihduntaa ja mutaatioita.

genetic code *geneettinen koodi* Kaava, jonka mukaan lähettiRNA:n emäskolmikot vastaavat translaation aloitusta, lopetusta tai tiettyjä aminohappoja. Tämän kaavan perusteella geeni ohjaa sitä vastaavan polypeptidin syntyä.

genetic map, linkage map *geneettinen kartta, kytchentäkartta* Kromosomikartta, jossa geenien tai geenimerkkien sijainti on esitetty suhteessa toisiinsa. Ei välttämättä vastaa täysin fyysistä sijaintia. Myös geenien sisäistä rakennetta kuvaavia karttoja kutsutaan geenikartoiksi.

genetic marker *geenimerkki* Geneettinen ero, jota käytetään hyväksi geneettistä rekombinaatiota tutkittaessa. Geenikarttaan liittyvä ominaisuus (esim. perinnöllinen sairaus), jota vastaavan geneettisen rakenteen sijainti kromosomissa tunnetaan. Katso marker.

genetic risk *geneettinen riski* Mahdollisuus jonkin ominaisuuden, yleensä sairauden, ilmenemiseen.

GeneWise Algoritmi/ohjelma, jota käytetään genomisten DNA-sekvenssien etsinnässä ja analyyseissä. Perustuu kätettyihin Markovin malleihin. <http://www.sanger.ac.uk/Software/Wise2>

genome *perimä, genomi* Eliön tai solun kromosomien sisältämä perinnöllinen informaatio.

genomic DNA *genominen DNA* Solun DNA-fraktio, joka muodostuu kromosomeista (ei siis sisällä soluelinten DNA:ta).

genotype *genotyyppi* Solun tai yksilön geneettinen kokonaisuus. Katso fenotyyppi.

germ cell *itusolu* Monisoluisen yksilön lisääntymissolu

germ line *itulinja, iturata* Sukusolujen syntyyn johtava solulinja.

global alignment *kokonaisrinnastus* Sekvenssien rinnastaminen niin, että sekvenssien samankaltaisuuksia pyritään löytämään koko pituudelta. Katso local alignment.

graph theory *graafiteoria, verkkoteoria* Graafiteoriaa voidaan käyttää tapauksissa, joissa tukittava ongelma voidaan esittää verkkomaisena kuvana, joka koostuu solmupisteistä eli kärjistä (vertex) sekä niitä yhdistävistä väleistä eli särmistä (edge). Bioinformatiikassa graafiteoriaa voidaan soveltaa mm. sekvenssoinnissa ja proteomiikassa esiin tuleviin ongelmiin, joissa ratkaisu kootaan suuresta joukosta osaratkaisuja (esim. sekvenssi sekvenssipaloista). Katso myös Eulerian path ja Hamiltonian path.

Greek key motif *Greek key -motiivi* β -säikeiden muodostama rakenteellinen motiivi.

GROMACS Biologisille makromolekyyleille kehitetty mallitusohjelmisto ja voimakenttä. <http://rugmd4.chem.rug.nl/~gmx/>

H

Hamiltonian path/cycle *Hamiltonin polku/ketju* Graafiteorian mukaisessa esityksessä reitti, joka kulkee graafin kaikkien solmupisteiden (vertex) kautta. Mikäli polku kulkee niin, että se lopuksi palaa lähtöpisteeseensä, kyseessä on Hamiltonin ketju. Katso graph theory.

haploid *haploidi* Tuma, solu tai yksilö, jossa on vain yksi kopio kutakin kromosomia, esim. sukusolut.

haplotype *haplotyyppi* Geeniryhmä, jossa kutakin geeniä vastaa yksi alleeli, jotka yhdessä muodostavat genotyypin. Myös kromosomi tai kromosomin osa, jonka alleelit ovat kytkeytyneet toisiinsa.

HCA (hydrophobic cluster analysis) katso hydrophobic cluster analysis.

helical wheel *kierrekaavio* Peptidisekvenssin esitystapa, jossa kuvataan sivuketjujen sijoittuminen α -kierteen eri puolille. Käytetään amfipaattisten kiertei-

den etsimisessä.

helix-coil transition *kierteen purkautuminen* Nukleinihappojen kaksoiskierteen tai proteiinien rakenteen purkautuminen esimerkiksi lämpötilan vaikutuksesta.

helix-loop-helix (HLH) motif *kierre-silmukka-kierre -motiivi* Kaksi α -kierrettä, joita erottaa silmukka. Transkriptiotekijöihin liittyvä motiivi, joka tunnistaa spesifisiä DNA-sekvenssejä ja sitoutuu niihin.

helix-turn-helix (HTH) motif *kierre-käännös-kierre -motiivi* Kaksi amfipaattista α -kierrettä, joita erottaa lyhyt β -käännös. Transkriptiotekijöihin liittyvä motiivi, joka tunnistaa spesifisiä DNA-sekvenssejä ja sitoutuu niihin.

heteroduplex *heterodupleksi* DNA-molekyyli, jossa juosteet ovat erilaisia. Juosteilla voi olla eri alkuperä, jolloin komplementaariset juosteet on saatu yhdistymään in vitro tai dupleksi muodostuu lähettiRNA:sta sekä vastaavasta DNA:sta.

heteroduplex analysis (mismatch analysis) *heterodupleksianalyysi* Eri alkuperää olevien DNA-juosteiden annetaan liittyä toisiinsa heterodupleksiksi, minkä jälkeen erot kartoitetaan. Kartoitusten menetelmiä on useita.

heterozygosity *heterotsygoottisuus* Tila, jossa vähintään yhden lokuksen alleelit ovat yksilön vastinkromosomeissa keskenään erilaiset.

heuristic *heuristinen, kokemusperäinen* Algoritmien yhteydessä heuristisella menetelmällä tarkoitetaan kokemusperäistä ja/tai oletuksiin perustuvaa menetelmää.

high throughput genome (HTG) Viimeistelemätöntä sekvenssiaineistoa, josta kuitenkin on hyötyä esimerkiksi sekvenssihauissa. Tietokannassa vähintään kahden kiloemäksen palasina. Nopeasti tuotettua raaka-aineistoa, jota ei ole varmennettu.

histone *histoni* DNA:han sitoutuva, sitä stabiloiva ja sen toimintaa säätelevä proteiini. Histonit jaotellaan viiteen eri tyyppi-

piin: H1, H2a, H2b, H3 ja H4. Histoniin avulla aitotumallisten (eukaryoottien) DNA pakkautuu tiiviisti.

HMM (Hidden Markov Model)*kätetty Markovin malli* Malli, jossa havaintoja mallitetaan piilevällä kerroksella, jolle on määritelty Markov-ominaisuus (mallin seuraava tila riippuu vain edellisestä tilasta).

HMMer Ohjelmapaketti, joka käyttää HMM-menetelmää sekvenssianalytiikassa. <http://hmmer.wustl.edu/>

hnRNA (heterogeneous nuclear RNA) *heterogeeninen tuman RNA* Lyhytikäinen ja suurimolekyylinen lähettiRNA:n esiaste.

homeobox *homeolaatikko* Konservoitunut DNA-jakso, joka alunperin löytyi banaanikärpäsen useista homeoottisista tai segmentaatiomutaatioita aiheuttavista geneistä.

homeodomain *homeodomeeni* Homeolaatikko-sekvenssin koodaama proteiinin osa.

homologous recombination *homologinen rekombinaatio* Tekijäinvaihdunta homologisten lokusten välillä.

homologue *homologi, samankaltainen sekvenssi* Evolutiivinen sukulaissekvenssi.

homology *homologia* Evolutiivinen samankaltaisuus, joka perustuu yhteiseen kantamuotoon.

homozygosity *homotsygoottisuus* Tila, jossa lokusten alleelit vastinkromosomeissa ovat samanlaiset. Katso heterozygosity.

Hopp-Woods Yksi proteiinien hydropaattisuuden ennustusmenetelmistä.

hormone response element (HRE) *hormonivastealue* DNA-sekvenssi, jonka hormonireseptori tunnistaa ja jonka ilmenemistä reseptori säätelee.

hot spot DNA-alue, jossa muita alueita selvästi useammin tapahtuu evoluutiivisia muutoksia, esim. mutaatioita.

HRE (hormone response element) katso hormone response element.

HTG (high throughput genome) katso high throughput genome.

HTML (HyperText Markup Language.) WWW-sivujen kuvauskieli.

HTTP (HyperText Transfer Protocol.) WWW-palvelimien käyttämä yhteydenpitomenetelmä.

hybridization *yhdistyminen, hybridisaatio*
1. Tapahtuma, jossa kaksi vastakkaista nukleiinihappojuostetta liittyy yhteen. Käytetään geeniteknikassa nukleiinihappojen tunnistusmenetelmänä. Voidaan käyttää diagnostiikassa tiettyjen DNA-jaksojen osoittamiseen näytteestä (katso: probe, koetin). 2. Risteytyminen. 3. Solujen yhdistäminen keinotekoisesti.

hydrogen bond (H-bond) *vetysidos* Elektronegatiivisen atomin ja toiseen elektronegatiiviseen atomiin kovalenttisesti sitoutuneen vetyatomin välinen vuorovaikutus (esimerkiksi C=O...H-N). Vetysidokset ovat voimakkaimpia biomolekyyleissä esiintyvä ei-kovalenttisia vuorovaikutuksia. Niillä on merkittävä vaikutus mm. emäsparien muodostumiseen ja proteiinien laskostumiseen.

hydropathy *hydropaattisuus* Vesihakuisuuden ja -pakoisuuden yleistermi.

hydropathy plot *proteiinin hydropaattisuuden ennuste* Kuvaaja, jossa on esitetty proteiinisekvenssin vesipakoisuus tai vesihakuisuus sekvenssin funktiona. Yleensä arvo lasketaan käyttämällä tietyn mitausta tarkastelualueita ja (painotettua) keskiarvoa, jolloin myös sekvenssissä tutkittavan aminohapon ympärillä olevat kohdat tulevat huomioiduiksi.

hydrophilicity *vesihakuisuus, hydrofiilisyyys*

hydrophilic residue *vesihakuinen aminohappo*

hydrophobic cluster analysis (HCA) *hydrofobinen klusterianalyysi* Ennustaa hydropaattisuutta ja sekundäärirakenteita.

hydrophobic interaction *hydrofobinen vuorovaikutus* Molekyylien välinen näennäinen vuorovaikutus, joka aiheutuu ni-

iden vettä hylkivästä luonteesta.

hydrophobicity *vesipakoisuus, hydrofobisuus* Termi, joka kuvaa ei-polaaristen molekyylien taipumusta pyrkiä pois vesiympäristöstä. Aminohappojen sivuketjujen vesipakoisuus vaihtelee suuresti, eikä sen kuvaamiseen ole yksiselitteistä keinoa tai asteikkoa. Aminohapoille yleisesti käytettyjä hydrofobisuusasteikkoja ovat mm. Kyte-Doolittle ja Hopp-Woods.

hydrophobic moment *hydrofobinen momentti* Molekyyliin tai sen osaan vaikuttavien vesihakuisten ja vesipakoisten vuorovaikutusten summaa kuvaava suure. Vaikuttaa merkittävästi esimerkiksi kalvoproteiineissa.

hydrophobic zipper (leucin zipper)

hydrofobinen vetoketju DNA:han sitoutuvissa proteiineissa esiintyvä rakenne, jolla proteiini tunnistaa sitoutumiskohdan DNA:ssa. Rakenne muodostuu kahdesta rinnakkaisesta kierteestä, joissa molemmissa joka seitsemäs aminopoppo on vesipakoinen (yleensä leusiini).

I

inborn error *synnynnäinen virhe*

indel (insertion or deletion) lisäys tai poistuma DNA- tai proteiinimolekyyllisissä.

information theory *informaatioteoria* Todennäköisyysteorian osa-alue.

in frame *lukukehyksessä, lukukehyksen säilyttävää*

initiation codon *aloituskodoni* lähettiRNA:n emäskolmikko (yleensä 5'-AUG), josta proteiinisynteesi alkaa.

initiator tRNA *aloitus-tRNA* Aminohapposynteesin aloittava tRNA.

insertion *lisäys* Yhden tai useamman DNA-nukleotidin tai aminohapon lisäys. Mutaatiotyyppi.

InsightII Kaupallinen molekyyylimallitusohjelma, joka sisältää useita biologisten makromolekyylien mallitukseen

soveltuvia ominaisuuksia. <http://www.accelrys.com/insight/index.html>

in silico tietokoneessa Toimenpide joka tehdään laskennallisesti. Esim. erilaiset simulaatiot.

in situ paikan päällä, paikallaan Tekniikat, joissa tutkittava solu, kudokset tai solukko on joko alkuperäisellä paikallaan tai kiinnitetty esim. lasilevyille.

Intelligenetics DNA- ja proteiinisekvenssi-analyysityökaluja kehittävä ja myyvä yritys.

intron introni Geeninsisäinen exonien välissä oleva ei-ilmentyvä DNA-alue. Introni kopioidaan RNA:ksi, mutta poistetaan siitä ennen proteiinisynteesiä. Katso: exon ja splicing.

inverse PCR käänteinen PCR

inversion kääntymä Kromosomin osa irtoaa alkuperäiseltä paikaltaan, kääntyy 180 astetta ja liittyy takaisin kromosomiin. Mutaatiotyyppi

inverted repeat (IR)käänteistoisto Nukleotidisekvenssi, joka on käänteinen verrattavalle sekvenssille. Useimmiten alle 50 emästä. IS-elementtien osia. (IS-elementti, insertion sequence, on transposonityyppi).

in vitro koeputkessa Toimenpide, joka tehdään keinotekoisissa olosuhteissa, esim. koeputkessa tai kasvatusalustalla.

in vivo elimistössä tai solussa Toimenpide, joka tehdään luonnollisissa olosuhteissa, esimerkiksi elävässä solussa tai kudoksessa.

IR (inverted repeat) katso inverted repeat.

isoschizomer isoskitsomeeri Restriktioenstyymit, jotka tunnistavat saman sekvenssin, ovat keskenään isoskitsomeereja.

J

Jukes-Cantor model Jukesin ja Cantorin malli DNA:n emäskorvautumismalli, jossa kaikki korvautumismahdollisuudet ovat yhtä todennäköisiä.

K

karyotype karyotyyppe Solun tai organisin kaikki kromosomit, erityisesti mitotistista soluista valokuvattujen ja pareittain rakenteen tai lukumäärän perusteella järjestettyjen kromosomien kokonaisuus.

kb (kilobase) katso kilobase (kb), kilobasepair (kbp).

kbp (kilobasepair) katso kilobase (kb), kilobasepair (kbp).

KEGG Kyoto Encyclopedia of Genes and Genomes Palvelu johon on koottu tietoa molekyylibiologisista reaktio- ja signaaliereiteistä sekä molekyylien ja geenien välisistä vuorovaikutuksista. <http://www.genome.ad.jp/kegg/>

kilobase, kilobasepair (kb, kbp)kiloemäs, kiloemäspari Tuhannen nukleotidin pituinen DNA- tai RNA-jakso.

Kimura model Kimuran malli 2-parametrinen ja 3-parametrinen Kimuran malli ovat yleisesti käytettyjä DNA:n emäskorvautumismalleja. 2-parametrisessa mallissa transitioiden todennäköisyys on eri kuin transversioiden todennäköisyys. 3-parametrisessa mallissa mallia on kehitetty siten, että eri transversioille käytetään kahta eri parametria.

knock-out poistogeeninen Eliö, jonka joku geeni on poistettu toiminnasta.

Kozak sequence Proteiinisynteesin lähetäRNA-aloituskodonin koodaaman aminohapon ympäriltä valmiista proteiinista löytyvä konservoitunut jakso.

ktup Monien hakualgoritmien (mm. FastA) parametri, joka määrittelee, kuinka pitkiä sekvenssialueita haun ensimmäisessä vaiheessa tarkastellaan. Parametri vaikuttaa sekvenssihaun herkkyyteen ja nopeuteen (arvo on tyypillisesti 1 tai 2 proteiineille, 3-6 nukleinihapoille; mitä pienempi, sitä herkempi, mutta hitaampi).

Kyte-Doolittle Algoritmi aminohaposekvenssin hydropaattisuuden laskemiseen.

L

lagging strand *viivästynyt ketju, jälkijuoste* DNA:n kahdentuessa toinen syntyvistä juosteista. Syntyy pieninä, myöhemmin yhteenliitettävänä palasina.

LCR (locus control region) katso locus control region.

leader sequence *johtosekvenssi* Esitumallisten eliöiden aminohapposynteesin säätelyyn liittyvä sekvenssi. Johtosekvenssi koodaa johtopeptidiä, joka sisältää useita kappaleita sitä aminohappoa, jonka synteesin säätelyyn kyseinen johtosekvenssi liittyy.

leading strand *johtava ketju, johtajuoste* DNA:n kahdentuessa uusista juosteista se, joka syntyy jatkuvana kokonaisuutena.

LINE (long interspersed repeat element) katso long interspersed repeat element.

linkage *kytkentä* Eri lokusten alleelit voivat periytyä yhdessä, jos ne sijaitsevat samassa kromosomissa. Etäisyyden mittana käytetään lokusten välistä rekombinaatio-taajuutta. Kytkentää hyödynnetään esim. geenikarttojen teossa.

Linkage ohjelma kytkentäanalyysiin. Ohjelma on käyttökelpoinen suurille sukupuille, mutta laskentalgoritmi mahdollistaa vain muutaman merkkigeenin samanaikaisen käytön. <http://linkage.rockefeller.edu/soft/linkage/>

local alignment *paikallinen rinnastus* Vertailtavista sekvensseistä etsitään huomattavan samankaltaisia alueita. Tulos esitetään yleensä taulukkona, jossa sekvenssien vastinkohdat ovat samassa sarakkeessa. Vert. kokonaisrinnastus, global alignment.

locus (pl. loci) *lokus* Geenin paikka kromosomissa tai kromosomikartassa.

locus control region (LCR) Alue, jolla geenin säätely ohjaavat elementit (promootorit ja tehostajat) sijaitsevat.

lod score Kymmenkantainen logarit-

mi kytkentää testaavasta uskottavuusosamäärästä. Kahden lokuksen kytkentää esittävä tilastotieteellinen arvo. Jos arvo on kolme tai suurempi, katsotaan lokusten olevan kytkeytyneitä toisiinsa.

Logo Usean sekvenssin rinnastusohjelma.

LOH (loss of heterozygosity) *heterotsygoti-avaje* katso loss of heterozygosity.

long interspersed repeat element (LINE) Usean kiloemäksen pituinen nisäkkäiden genominen toistojakso.

long terminal repeat (LTR) Usean sadan emäksen pituiset käänteiset toistojaksot transposonien ja virusten esiasteiden DNA-ketjun päissä. Vaikuttavat transposonien ja viruksien esiasteiden liittymiseen isäntä-DNA:han.

loop *silmukka* Yleisnimitys lyhyille proteiini- tai nukleiinihapporakenteille, jotka yhdistävät sekundäärirakenteita. Atk: ohjelmassa oleva toistettava osuus.

loss of heterozygosity (LOH) Näytteiden välinen ero heterotsygoottien määrässä. Jos toinen näyte on selvästi homotsygoottisempi, kutsutaan ilmiötä LOH:ksi.

LTR (long terminal repeat) katso long terminal repeat.

M

MACAW Ohjelma, jolla voi paikallisesti rinnastaa sekvenssejä.

main chain *pääketju, selkäranka* Monomeerit toisiinsa liittävä polymeerin (esimerkiksi proteiinin tai nukleiinihapon) rakenneosa.

marker *merkkijakso* Genominen tai soluelimen DNA:n emäsjärjestys, joka eroaa tarpeeksi yksilöiden välillä, jotta sen periytymistä sukulinjassa ja/tai erilaisissa soluissa voidaan jäljittää. Merkkijakso voi olla geenissä tai geenien ulkopuolella.

masking *peittäminen, suodatus*. poistaa yksinkertaiset toistot ja homopolymeeriset jaksot sekvenssistä ennen tietokantahakua tai linjausta, jotta ne eivät antaisi väärää positiivisia tuloksia ja häirit-

sisi hakua. Suodatusohjelmien (esim. SEG proteiineille ja DUST nukleiinihapoille) peittämät aminohapot esitetään usein rinnastuksissa X-kirjaimina ja emäkset N-kirjaimina

maximum likelihood suurimman uskottavuuden menetelmä Menetelmä käytetään mm. evolutiikassa etsittäessä fylogeneettistä puuta, joka parhaiten toteuttaa asetetun evoluutiomallin.

maximum parsimony parsimoniamenetelmä, niukkuusmenetelmä Sukulaissuhteiden selvitysmenetelmä, jonka tulostama fylogeneettinen puu vähimmin mahdollisin muutoksien selittää aineiston.

Mb (Megabase, Megabyte) katso Megabase ja Megabyte.

ME (Minimum Evolution) katso Minimum Evolution.

Megabase (Mb) Miljoona emästä. Nukleotidisekvenssin pituuden mittayksikkö.

Megabasepair (Mbp) *megaemäs, megaemäspari* Miljoona emästä pitkä nukleiinihappojakso.

Megabyte (MB) *megatavu* Tiedoston tai tietokonelevyn koon tai kapasiteetin mitta. Yksi megatavu sisältää 2 potenssiin 20 tavua eli noin miljoona tavua.

metabolome *metabolomi* Eliön tai solun aineenvaihdunnan kokonaisuus.

methylation metylaatio Metyyliryhmien (CH₃) liittäminen molekyyliin. Metylaatio voi kohdistua esimerkiksi genomisen DNA:n tiettyihin nukleotideihin, mikä säätelee geenien transkriptiota.

Metropolis algorithm *Metropolis-algoritmi* Satunnaisuutta käyttävä optimointitehtävien ratkaisumenetelmä.

microsatellite mikrosatelliitti Genomissa runsaita yhdestä viiteen emäksen pituisia toistojaksoja, joiden lukumäärä vaihtelee yksilöittäin. Voidaan käyttää merkkijaksoina.

midnight zone pimeä alue Sekvenssejä vertaillaessa alue, jolla vertailun luotettavuutta ei voida arvioida.

Minimum Evolution (ME) *minimievoluutio* Menetelmä, joka etsii lyhimmän etäisyys ehdot täyttävän puukaavion.

minisatellite minisatelliitti Noin 10-100 emästä pitkiä toistojaksoja, joiden määrä vaihtelee yksilöittäin. Yleensä lähellä kromosomin päätä. Voidaan käyttää merkkijaksona.

minus-strand miinus-juosteinen (-nauhainen, säikeinen) RNA-virusten komplementaarinen DNA-juoste. Katso (-)strand.

mismatch repair yhteensopimattomuuden korjaaminen DNA-synteesin aikana väärin pariutuneiden nukleotidien etsiminen ja korvaaminen.

missense mutation Mutaatio, joka muuttaa kodonin nukleotidijärjestystä niin, että sen koodaama aminohappo muuttuu.

mitochondrio mitokondrio Aitotumallisten solujen soluelin, jossa soluhengitys tapahtuu. Mitokondriot sisältävät rengasmaisia DNA-molekyylejä, ns. mitokondriogénomien.

molecular clock molekyylikello Hypoteesi, jonka mukaan molekyylien evoluutio on ajan suhteen vakio eli molekyylien eroista voidaan suoraan laskea niiden erkanemisajankohta.

molecular dynamics molekyyliidynamiikka Simulointimenetelmä, jossa molekyylien liikettä seurataan ajan funktiona. Yleensä laskennassa käytetään klassisen mekaniikan liikeyhtälöitä.

molecular mechanics molekyylimekaniikka Menetelmä molekyylien rakenteen ja dynamiikan laskemiseen. Yleisnimitys laskennallisille menetelmille, joissa tutkitaan molekyylin avaruusrakenteen muutoksien vaikutusta potentiaalienergiaan. Yleisimmin käytetty molekyylimekaaninen tehtävä on voimakenttään perustuva komiulotteisen rakenteen optimointi.

molecular modeling molekyyli mallitus Molekyylin avaruusrakenteen tutkimus tai ennustus rakennemallien avulla. Tarkoittaa yleensä tietokoneavusteista

molekyylien kolmiulotteisten mallien käsittelyä.

monocistronic mRNA lähettiRNA, joka translaatiossa tuottaa vain yhden polypeptidiketjun.

monophyletic *monofyleettinen* Ominaisuus tai seikka, joka liittyy sellaiseen populaation sisäiseen periytymiseen, johon ulkopuoliset populaatiot eivät ole vaikuttaneet.

Monte Carlo method *Monte Carlo -menetelmä* Satunnaisuuteen perustuva numeerinen menetelmä.

mosaicism *mosaiikkisuus* Tila, jossa yksilöllä on vähintään kaksi erilaista solulinjaa t.s. yksilön kaikki solut eivät ole samaa genotyyppiä. Aiheutuu geneettisen informaation muuttumisesta tsygotin muodostumisen jälkeen. Voidaan saada aikaan myös keinotekoisesti yhdistämällä kahden saman eliölajin yksilön soluja.

most parsimonious tree (MPT) parsimonisin puu. Fylogeneettinen puu, joka selittää aineiston yksiköiden väliset erot vähimmin mahdollisin muutoksina.

motif *motiivi, aihe* Lyhyehkö jollekin piirteelle ominainen sekvenssijakso tai rakennekomponentti.

MPT (most parsimonious tree) katso most parsimonious tree.

mRNA (messenger RNA) *lähetti-RNA* Geenin DNA-juosteelle komplementaarinen yksijuosteinen RNA-molekyyli, joka syntyy transkriptiossa. LähettiRNA ohjaa proteiinisynteesiä ja määrää polypeptidin aminohappojärjestyksen.

MultiAlin Usean sekvenssin rinnastusmenetelmä. <http://www.toulouse.inra.fr/multalin.html>

MultiDisp Usean sekvenssin rinnastuksen visualisointiohjelma. <http://protein.uta.fi:8080/cgi-bin/MultiDisp.cgi>

multifactorial disease *monitekijäinen sairaus* Tila, jonka syntyyn vaikuttaa vähintään kaksi geeniä ja ympäristö. Käytetään myös termiä polygenic disorders. Katso multigenic disorder or disease

ja complex disease sekä polygenic disorders.

multigenic disorder or disease *monigeeninen häiriötila tai sairaus* Tila, jonka syntyyn vaikuttaa vähintään kaksi geeniä ja ympäristö. Käytetään myös termiä polygenic disorders. Katso multifactorial ja complex disease sekä polygenic disorders.

multiple sequence alignment (MSA) *monien sekvenssin rinnastus* Sekvenssiaineiston esittäminen niin, että samankaltaiset kohdat on aseteltu päällekkäin.

multiplexing *limitys* Usean yhdistetyn näytteen yhtäaikaista sekvenssointia. Myös atk-termi.

multipoint linkage analysis *monipistekyt-kentäanalyysi* Kytkentäanalyysi, jossa uuden lokuksen sijaintia kytkentäryhmässä (kromosomissa) tutkitaan usean sellaisen lokuksen avulla, joiden suhteelliset etäisyydet toisiinsa tiedetään.

mutagenesis *mutageneesi* Mutaation aiheuttaminen DNA:han esimerkiksi säteilyttämällä, kemikaalilla tai transposonilla.

mutation *mutaatio, muuntuma* Useimmiten rajoitettu tarkoittamaan organismin DNA:n emäsjärjestyksen muuttumista ilman rekombinaatiota.

N

NCBI (National Center for Biotechnology Information) Yhdysvaltalainen keskus, jolla on mm. avoin palvelin biotieteellisen aineiston etsimiseen, varastointiin ja analysointiin. <http://www.ncbi.nlm.nih.gov/>

NDB (Nucleic Acid Database) Nukleiinihapporakenteita sisältävä tietokanta. <http://ndbserver.rutgers.edu/>

Needleman-Wunsch algorithm *Needlemanin ja Wunschinin algoritmi* Dynaamiseen ohjelmointiin perustuva sekvenssivertailualgoritmi kahden sekvenssin rinnastukseen. Algoritmi etsii kahden sekvenssin välisen parhaan mahdollisen kokonaisrinnastuk-

sen.

neighbor-joining algorithm *neighbor joining -algoritmi, NJ-algoritmi* Evolutiikassa: puukaavioiden luomisessa käytettävä heuristinen menetelmä.

neural network (artificial neural network) *neuroverkko, hermoverkko* Algoritmi, jonka toiminta jäljittelee tietojenkäsittelyä biologisissa hermoverkoissa. Neuroverkkoalgoritmien tarkka määritelmä vaihtelee eri yhteyksissä. Yleisiä tunnusmerkkejä ovat mm. 1. Tieto syötetään neuroverkkoon oppimisprosessin avulla, 2. verkko sisältää useita toisiinsa kytkeytyneitä yksikköjä, joiden välisiin vuorovaikutussuhteisiin tieto varastoituu.

non-coding strand *koodaamaton juoste* Geenille komplementaarinen DNA-juoste.

nonparametric analysis *ei-parametrinen analyysi* Tilastotieteessä: ei oleteta mitään aineistoa kuvaavien jakaumien muodosta.

nonsense mutation DNA:n emäsjärjestyksen muutos, jonka seurauksena proteiinisynteesi pysähtyy. Stop-kodonin tuottava mutaatio.

nonsynonymous substitution *ei-synonyminen substituutio* Aminohapon korvaus täysin toisentyypisellä aminohapolla.

NP-hard (Nondeterministic polynomial complete/hard) *NP-täydellinen tai -vaikea* Laskentaongelman ominaisuus. NP-täydellisille ongelmille tunnetaan vain hyvin hitaita ratkaisualgoritmeja, joiden aikavaatimus kasvaa eksponentiaalisesti algoritmin syöttötietojen määrän mukana (ns. kombinatorinen räjähdys). Tämän takia NP-täydellisten ongelmien ratkaisuun käytetään usein (epätarkkoja) heuristisia menetelmiä.

nuclear localization signal *tumaanohjaussignaali* Aitotumallisissa tumaan kuljetettävien proteiinien signaalina toimiva peptidisekvenssi.

nucleic acid *nukleiinihappo* DNA ja RNA ovat nukleotideista rakentuvia nukleiinihappoja.

nucleoside *nukleosidi* Nukleiinihapon ja nukleotidin osa. Puriini- tai pyrimidiiniemäs + pentoosisokeri.

nucleosome *nukleosomi* Kromatiinin kromosomiksijärjestymisen perusyksikkö. Koostuu n. 200 emäsparista ja histoniproteiiniytimestä.

nucleotide *nukleotidi* Nukleosidi + fosfaatti. Nukleiinihapon rakenneyksikkö.

nucleotide-binding domain nukleotidin sitova domeeni proteiinissa.

O

oligonucleotide *oligonukleotidi* Korkeintaan muutamia kymmeniä emäksiä sisältävä DNA-sekvenssi.

OMIM (Online Mendelian Inheritance in Man) Geneettisten sairauksien tietokanta. <http://www.ncbi.nlm.nih.gov/Omim/>

oncogene *onkogeeni, syöpägeeni* Muta-toitunut ja/tai yli-ilmennetty normaaligeenin versio eläinsolussa tai syöpää aiheuttavassa viruksessa. Sekoittaa solun normaalit kasvutoiminnot ja joko yksin tai yhdessä muiden muutosten kanssa saa aikaan pahanlaatuisen kasvaimen kehittymisen.

open reading frame (ORF) *avoin lukukehys* Mahdollisesti proteiiniksi käännettävä DNA-sekvenssi, joka alkaa aloituskodonilla ja päättyy lopetuskodoniin. Katso: lukukehys, reading frame

operational taxonomic unit (OTU) *taksoni*

operator *operaattori* DNA-alue, johon sitoutumalla estäjäproteiini estää transkription aloittamisen lähellä sijaitsevasta promoottorista.

operon *operoni* Bakteerien geenitoiminnan yksikkö, johon kuuluu promoottori, operaattori, sekä yksi tai useampia rakennegeenejä.

ORF (open reading frame) katso open reading frame.

orthology *ortologia, ortologisuus* Lajiutumisen seurauksena syntynyt

samankaltaisuus.

OTU (operational taxonomic unit) *taksoni outgroup* *ulkoryhmä* Taksoni, jonka ei katsota kuuluvan fylogeneettisesti samaan ryhmään muiden vertailtavien taksoneiden kanssa.

overlapping genes *päällekkäiset geenit* Erilisiä geenejä, joiden ilmennettävät alueet ovat osittain päällekkäiset ja joita luetaan eri lukukehyksessä.

P

palindrome *palindromi, käänteinen toistojakso* DNA-jakso, joka on sama luettuna kaksoskierteen kummastakin juosteesta. Esimerkiksi esitumallisissa lähellä transkription lopetuskohtaa sekä useimpien restriktioentsyymien tunnustussekvensseissä.

PAM (Point Accepted Mutation). Sarja pisteytysmatriiseja, joilla määritetään aminohappoja vertailtaessa, miten aminohapot vastaavat toisiaan eli mikä on niiden samankaltaisuuden aste.

paralogy *paralogia, paralogi* Samankaltaisuus, joka johtuu geenin kahdentumisesta eli samankaltaisilla proteiineilla on samassa eliössä eri tehtävä.

parametric analysis *parametrinen analyysi* Tilastollinen menetelmä, joka tekee oletuksia populaation jakaumasta tarkastellun ominaisuuden suhteen.

parsimony katso maximum parsimony.

PAUP (Phylogenetic Analysis Using Parsimony) Fylogeniikkaohjelmisto, jonka avulla voidaan analysoida sekvenssien ja taksonien sukulaisuussuhteita useilla eri menetelmillä. <http://paup.csit.fsu.edu/>

PCR (polymerase chain reaction) katso polymerase chain reaction.

PDB (Protein DataBank.) Makromolekyylien kolmiulotteisten rakenteiden tietokanta. <http://www.rcsb.org>

PDBSum PDB-tietokannasta johdettu tietokanta, jonka tietueet sisältävät rak-

enteen lisäksi yhteenvedon rakenteesta sekä luokitteluja ja analyysituloksia. <http://www.biochem.ucl.ac.uk/bsm/pdbsum/index.html>

penetrance *ilmenevyys, läpäisevyys* Ehdollinen todennäköisyys ilmiäiselle kun genotyyppi on annettu. Esim. merkintä $P(\text{sairas} | \text{aa}) = 0.8$ tarkoittaa, että todennäköisyys sairastua on 80 prosenttia, jos kantaa genotyyppiä aa. Penetranssitodennäköisyydet voivat olla erilaisia mm. iän tai sukupuolen suhteen.

peptide *peptidi* Lyhyt aminohappoketju, pieni proteiini tai proteiinin pilkottu osa.

peptide bond *peptidisidos* Kovalenttinen sidos, jolla aminohapon pääketjun karboksyyli-ryhmä liittyy toisen aminohapon aminoryhmään.

Pfam proteiiniperhetietokanta. Proteiinit on jaoteltu sekvenssin mukaisesti perheisiin. Kullekin perheelle tietokanta sisältää joukon hyvin rinnastettuja sekvenssejä, joiden perusteella luodulla HMM-mallilla muita sekvenssejä voidaan verrata ja luokitella. <http://www.sanger.ac.uk/Software/Pfam/>

phage display Faagikirjasto, jossa litännäiset jaksot ilmenevät osana faagin kuoriproteiinin pintarakennetta. Käytetään esimerkiksi vasta-ainepitooppien tunnistuksessa.

PHD Sekvenssiin perustuva proteiinirakenteen ennustusohjelmisto, joka ennustaa annetun sekvenssin sekundäärirakenteen, pintarakenteen sekä solukalvon läpäisevien sekvenssien sijainnin.

phenotype *ilmiasu* Yksilön ominaisuuksien kokonaisuus, joka syntyy genotyypin ja ympäristön vuorovaikutuksesta.

phyla *pääjakso* Lajien luokittelun perusryhmä, jonka jäsenet ovat keskenään samankaltaisia ja oletettavasti sukulaisia.

Phylip Fylogeniikkaohjelmisto, jonka avulla voidaan analysoida sekvenssien ja taksonien sukulaisuussuhteita useilla eri menetelmillä. <http://evolution.genetics.washington.edu/phylip.html>

phylogeny *fylogenia* Taksonin tai muun ryhmän kehityshistoria.

phylogram *fylogrammi* Taksonin kehityshistoriaa kuvaava kaavio.

PIR (Protein Identification Resource) Sekvenssitietokannoista johdettu tietokanta, joka on tarkoitettu proteiinisekvenssien ja niitä vastaavien nukleinihapposekvenssien tunnistamiseen ja analysointiin. <http://www-nbrf.georgetown.edu/pir/>

pleated sheet *aaltoileva tai poimuinen levy* β -tason tyypillinen rakenne, jossa proteiinin pääketjun muoto aaltoilee. Yksittäisen β -nauhan perättäiset α -hiilet sijaitsevat vuorotellen β -tason keskikohdan eri puolilla ja muodostavat β -tasolle tyypillisen aaltoilevan hienorakenteen.

point mutation *pistemutaatio* Yhden nukleotidin muutos DNA-sekvenssissä. Käytetään yleisesti myös yhden aminohapon muutoksista.

polygenic (character) *polygeeninen ominaisuus* Polygeeniseen ominaisuuteen vaikuttaa useita eri geenejä, joiden yksittäiset vaikutukset ovat usein pieniä. Katso complex, multigenic ja multifactorial disease.

polymerase chain reaction (PCR) *polymerasiketjureaktio* Tehokas tapa monistaa DNA:ta in vitro. Perustuu lämpövaihteluita kestävään DNA-polymeraasiin.

polymorphism *monimuotoisuus, vaihtelu, polymorfismi, polymorfia* Populaatiossa on kaksi tai useampia ilmiäsuja (saman geenin suhteen), joita ei voi selittää pelkästään mutaatioilla.

post-transcriptional modification *transkription jälkeinen muokkaaminen* Transkriptiossa syntyneen RNA:n rakenteeseen ennen translaatiota tehtävät muutokset. Esimerkiksi intronialueiden poisto, silmukointi (pujonta).

post-translational modification *translaation jälkeinen muokkaaminen* Translaatiossa tuotettuun aminohappoketjuun tehtävät muutokset, esim. signaalisekvenssien

poistaminen tai metyyli-, fosfaatti- ja sokeriryhmien liittäminen kovalenttisesti proteiiniin.

precursor *edeltäjä, prekursori* Aine tai seikka, joka edeltää seuraavaa vaihetta tai seikkaa. Käytetään esim. proteiinista, johon ei ole tehty aktiivisuuden synnyttämiseen tarvittavia muutoksia. Nukleinihappojen yhteydessä: fosforyloitu nukleinihappo, joista nukleinihappoketju syntetisoidaan.

Pribnow box *Pribnow-sekvenssi* DNA-sekvenssijakso (TATAAT), joka ilmaisee bakteerien DNA:ssa RNA-synteesin aloituskohdan. Sijaitsee 5'-päässä, 10 emästä ennen synteesin aloituskohtaa. Voi esiintyä muuntuneena.

primary structure *primäärirakenne, sekvenssi* Proteineissa: peptidisidosten yhdistämien aminohappojen järjestys proteiinissa tai peptidissä.

primer *aluke* Lyhyt polynukleotidiketju, jota käytetään käynnistämään DNA:n kahdentuminen.

proband Sukupuun tai perheen yksilö, jonka perusteella suku ja/tai perhe on valittu tutkimukseen.

probe *koetin* Molekyylibiologiassa: kemiallisesti leimattu molekyyli, esim. DNA- tai RNA-jakso, jota käytetään hybridisaatiossa tietyn sekvenssin etsimiseen. Bioinformatiikassa: sekvenssi, jolla tietokantahaku tehdään. Koettoimella viitatan yleensä tunnettuun sekvenssiin tai molekyyliin, jolla haku tehdään. Vrt. kohde (target)

ProDom Tietokanta, joka sisältää automaattisesti luokiteltuja proteiinidomeeniryhmiä. <http://www.toulouse.inra.fr/prodom.html>

profile alignment *profiilirinnastus* Sekvenssirinnastus, jossa pisteytysmatriisin sijasta tai lisäksi käytetään paikkakohtaista pisteytysmatriisia eli profiilia. Yleensä profiili tehdään tunnetun monen sekvenssin rinnastuksen perusteella. Profiilirinnastuksessa voidaan pisteytysmatriisin perustuvaa rinnastusta paremmin ot-

taa huomioon tietyille sekvenssiperheelle tyypilliset piirteet.

prokaryote (procaryote) *esitumallinen* Bakteerit ja syanobakteerit, joiden genomi (DNA-kromosomi) ei ole erillisen tumakalvon suojassa. Esitumallisissa soluissa on hyvin vähän soluelimiä.

promoter *promoottori* DNA-sekvenssin kohta, johon RNA-polymeraasi sitoutuu ennen transkription aloitusta. Esimerkiksi Pribnow- ja TATA-sekvenssit.

proofreading (editing) *oikoluku, tarkistus* Yleisnimitys DNA- ja proteiinisynteesissä vaikuttaville biokemiallisille reaktioille, joilla synteesissä tapahtuneita virheitä poistetaan.

Prosite Tietokanta, joka koostuu biologisesti merkittävistä sekvenssijaksoista ja -perheistä sekä niihin liittyvistä sekvenssimotiiveista ja -profiileista. <http://www.expasy.ch/prosite/>

proteome *proteomi* Eliön tai solun tuottamien proteiinien kokonaisuus.

pseudogene *pseudogeeni, valegeeni* Ilmentymätön DNA-jakso, jonka emäsjärjestys muistuttaa jonkin geenin emäsjärjestystä.

PSI-Blast (Positio Specific Iterative BLAST) BLAST-ohjelman versio, joka tekee BLAST-haun perusteella löytyneistä sekvensseistä sekvenssiprofiilin, jonka avulla ohjelma suorittaa uuden, tarkemman haun. Tulosten perusteella tehdään uusi sekvenssiprofiili ja haku niin kauan, kunnes merkittävästi samankaltaisia sekvenssejä ei enää löydy.

PubMed NCBI:n ylläpitämä bio- ja lääketieteen kirjallisuustietokanta. <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi>

puckering *taipuminen, rypistyminen* Molekyyliarakenteen taipuminen pois tasomaisesta rakenteesta.

purine *puriini* Tyypeä sisältävä kaksoisringsrakenteinen orgaaninen yhdiste. Toinen nukleiinihappojen emäsryhmistä. Puriinin sisältäviä nukleotideja ovat adeniini ja guaniini.

pyrimidine *pyrimidiini* Tyypeä ja aromaattisen rengasrakenteen sisältävä orgaaninen yhdiste. Toinen nukleiinihappojen emäsryhmistä. Pyrimidiinin sisältäviä nukleotideja ovat sytosiini, tymiini sekä urasiili.

Q

quartet-puzzling Heuristinen menetelmä, jonka avulla suurimman uskottavuuden menetelmää voidaan soveltaa suurillekin aineistoille. Menetelmässä aineistosta määritetään suurinta uskottavuutta vastaavat puut kaikille mahdollisille neljän taksonin ryhmille, joista parhaat yhdistetään.

quaternary structure *kvaternäärirakenne* Useammasta eri aminohappoketjusta koostuvan proteiinikompleksin kokonaisrakenne. Proteiinikompleksin alayksikköjen sijoittuminen toistensa suhteen.

query *hakusekvenssi* Sekvenssi, jolla tietokantahaku tehdään.

R

RACE (rapid amplification of cDNA ends) katso rapid amplification of cDNA ends.

Ramachandran plot *Ramachandran-kuvaaja, kiertokulmakartta, phi-psi-kuvaaja* Proteiinin aminohapot on esitetty kuvaajana, jossa x-akselilla on pääketjun phi-kulma ja y-akselilla vastaava psi-kulma. Kuvaajan avulla voidaan tarkastella proteiinin stereokemiaa ja sekundäärirakenteiden esiintymistä.

random coil *satunnaislaskostuma, satunnaisrakenne* Proteiinirakenteen alueet, joissa ei ole selvää sekundäärirakennetta. Käytetään myös laskostumattomasta tai denaturoidusta proteiini- tai nukleiinihapporakenteesta.

rapid amplification of cDNA ends (RACE) Tekniikka, jolla cDNA:n 5'- tai 3'-pää saadaan PCR-monistettua, kun osa päiden välisestä sekvenssistä on selvillä.

rare-cutter Restriktioentsyymi, jonka tunnistuskohtia on harvassa.

reading frame lukukehys Kodonien rajakohtien ja lukusuunnan määrittely DNA-sekvenssissä. Koska DNA-sekvenssi voidaan jakaa kolmen emäksen mittaiseksi kodonisarjaksi kolmella tavalla kahteen eri suuntaan, on jokaisella DNA-sekvenssillä 6 mahdollista lukukehystä, joista yksi on translaatiossa käytetty lukukehys.

recombination rekombinaatio, uudelleenyhdistely Tapahtuma, jossa perintötekijöiltään erilaiset (yhden tai useamman geenin suhteen) solut tai yksilöt tuottavat vanhempiin nähden uudentyyppisen jälkeläisen.

regulatory element säätelyosa Geenin ilmentymistä säätelevä (vähentävä tai voimistava) DNA-sekvenssin osa.

repeat toistojakso Sekvenssissä esiintyvä toistuva jakso.

replication replikaatio, kopioituminen Biologiassa: solun jakautumista edeltävä perimäaineksen (DNA:n) kahdentuminen.

replicon replikoni, replikaatioyksikkö Perimän osa, joka pystyy kopioitumaan itsenäisesti. Nukleiinihappoketjun osa, joka kopioituu saman replikaation aloituskohdan ohjaamana. Esitumallisilla on kromosomissa yksi replikoni, aiotumallisilla useampia.

residue tähte Polymerointireaktion jälkeen polymeeriin jäänyt monomeerin osa. (Nukleiinihappo- tai aminohappoketjun perusrakenneosa).

restriction endonuclease restriktioentsyymi, restriktioendonukleaasi Bakterofageja bakteereissa pilkkovia entsyymejä. Restriktioentsyymit tunnistavat tietyn, yleensä palindromisen, DNA-sekvenssikohtan ja katkaisevat kaksijuosteisen DNA:n tästä kohdasta. Käytetään yleisesti molekyylibiologisissa tekniikoissa.

restriction map restriktiokartta Kaavio, joka osoittaa restriktioentsyymien tunnistuskohtien sijainnin DNA:ssa.

reverse open reading frame (RORF) käänteinen avoin lukukehys.

reverse transcriptase käänteiskopioija tai -transkriptaasi Retrovirusten DNA-polymeraasientsyymi, joka pystyy tuottamaan DNA-sekvenssin RNA-sekvenssin perusteella. Käytetään yleisesti molekyylibiologisissa tekniikoissa.

rise nousu Esim. DNA:n tai proteiinin α -kierteen nousu. Ilmoitetaan yleensä muodossa: monomeeriä kierrosta kohden.

RNAMOT Tietokoneohjelma RNA:n primääri- ja sekundäärirakenteiden analyysiin sekä RNA-motiivien etsimiseen. <ftp://merck.bch.umontreal.ca/pub/Rnamot/>

RORF (reverse open reading frame) katso reverse open reading frame.

rRNA (ribosomal RNA) ribosomaalinen RNA Ribosomien rakenneosa. Suurin osa solun RNA:sta on erikokoisia ribosomaalisia RNA-molekyylejä.

RT-PCR (reverse transcriptase polymerase chain reaction) käänteistranskriptaasi polymeerasiketjureaktio Käänteiskopioijaentsyymiä käyttävä PCR.

S

salt bridge suolasilta Proteiinirakenteen sisäinen ionisidoksen kaltainen vuorovaikutus positiivisesti ja negatiivisesti varautuneen aminohapon välillä.

Sanger Institute The Wellcome Trust Sanger Institute Englantilainen genomitutkimuskeskus. <http://www.sanger.ac.uk/>

SAR (structure-activity relationship) katso structure-activity relationship.

satellite DNA satelliitti-DNA DNA-sekvenssin kohta, joka poikkeaa huomattavasti ympäristöstään yleensä tiheästi toistuvan sekvenssijakson johdosta ja jonka tehtävää ei tunneta.

SCOP (Structural Classification of Proteins) Tietokanta, jossa PDB:stä löytyvät proteiinirakenteet on ryhmitetty hierarkiseksi kokonaisuudeksi. <http://scop.mrc-lmb.cam.ac.uk/scop/>

scoring matrix *pisteytysmatriisi* Taulukko, jolla paikasta riippumattomissa sekvenssirinnastusmenetelmissä määritetään eri nukleotidi- tai aminohappoyhdistelmien vaikutus rinnastuksen laatua tai todennäköisyyttä kuvaavaan pisteytykseen. Katso PAM ja BLOSUM.

secondary structure *sekundäärirakenne* Proteiinissa paikallisesti esiintyvät säännölliset kolmiulotteiset rakenteet. Yleisimmät sekundäärirakenteet ovat α -kierre ja β -säie.

self-splicing *Itsesilmukointi* Silmukoiminen itsenäisesti ilman ulkopuolisten molekyylien apua.

sense-strand kaksinauhaisen DNA:n koodaava juoste. Tässä juosteessa nukleotidijärjestys on sama kuin lähettiRNA:ssa.

SeqLab GCG-ohjelmistopakettin X-yhteyden kautta käytettävä käyttöliittymä.

sequence *sekvenssi* Nukleotidi- tai aminohappojärjestys nukleiinihapoissa tai proteiineissa.

sequence-tagged site (STS) *sekvenssimerkitty alue* Muutaman sadan emäsparin pituinen DNA-sekvenssijakso, joka esiintyy genomissa vain yhdessä tunnetussa kohdassa ja on tunnistettavissa PCR-tekniikalla. Sekvenssitunnusalueita käytetään tunnistettaessa ja yhdisteltäessä eri lähteistä saatua sekvenssiaineistoa.

sequencing *sekvensointi* Nukleotidi- tai aminohappoketjun sekvenssin kokeellinen määrittäminen.

SeqWeb GCG-ohjelmistopakettin verkkosivustoin kautta käytettävä käyttöliittymä. Esim. CSC:n SeqWeb rekisteröidylle käyttäjille: <https://seqweb.csc.fi/gcg-bin/seqweb.cgi>

sex chromosome *sukupuolikromosomi* Kromosomi, jonka läsnäolo, periytyminen tai tietty muoto määrää yksilön sukupuolen.

Shine-Dalgarno sequence *Shine-Dalgarno-sekvenssi* BakteriilähettiRNA:ssa yleinen sekvenssijakso, jonka perusmuoto on 5'-AGGAGG-3'. Se tai samankaltainen sekvenssi esiintyy yleensä noin 7 emästä ennen aloituskodonia.

short interspersed repeat element (SINE) Lyhyt genominen toistojakso.

short tandem repeat (STR) lyhyt peräkkäinen toistojakso.

shotgun sequencing *satunnaissekvensointi (haulikkosekvensointi)* Menetelmä, jossa genomi pilkotaan satunnaisen pituisiksi paloiksi, joiden (kaikkien tai osan) emäsjärjestys määritetään.

signal sequence *signaalisekvenssi* Proteiinisekvenssin N-päässä oleva osa, joka avustaa solusta erittyvien proteiinien tai solukalvoproteiinien kulkeutumista solukalvolle ja/tai sen läpi. Yleensä signaalisekvenssi poistetaan synteessin tai siirtymisen jälkeen. Soluorganelleissa signaalisekvenssit saattavat sijaita ketjun keskellä, eikä niitä välttämättä poisteta proteolyttisesti.

silent mutation *hiljainen mutaatio, peittyvä mutaatio* DNA-muutos, joka ei näy geenin koodaaman proteiinin sekvenssissä.

Simwalk Ohjelma sekä parametriseen että ei-parametriseen kytkentäanalyysiin. Toisin kuin Genehunteria, ohjelmaa voi käyttää myös suurille sukupuille.

SINE (short interspersed repeat element) katso short interspersed repeat element.

single-gene disorder Tauti tai tila, jonka aiheuttaa yksi geeni.

Single Nucleotide Polymorphism (SNP) *yhden emäksen monimuotoisuus* Lajin sisäinen nukleotidisekvenssin tietyssä kohdassa esiintyvä vaihtelu.

single-strand conformational polymorphism (SSCP) Pistemutaatioiden havainnointitekniikka. Tietty sekvenssialue PCR-monistetaan ja ajetaan geelissä. Mutaatio voi muuttaa rakennetta, mikä muuttaa molekyylin liikkuvuutta geelissä.

site-directed mutagenesis *kohdennettu mutageneesi* Yleisnimitys laboratoriomenetelmille, joilla DNA-sekvenssiin tehdään haluttuun kohtaan yksittäisen tai

- useamman emäksen muutos.
- site-specific mutagenesis** katso site-directed mutagenesis.
- sliding window** *liukuva ikkuna* Sekvenssi-analyysissä: tarkastelualue. Sekvenssiä käydään läpi ikkunan pituisissa palasissa niin, että ensimmäinen ikkuna alkaa ensimmäisestä sekvenssikohdasta ja analyysin edetessä ikkunaa liu'utetaan kohta kerrallaan eteenpäin sekvenssissä.
- Smith-Waterman algorithm** *Smithin ja Watermanin algoritmi* Dynaamiseen ohjelmointiin perustuva sekvenssien rinnastusmenetelmä. Algoritmi etsii kahden sekvenssin väliset parhaat mahdolliset paikalliset rinnastukset.
- SNP** (single nucleotide polymorphism) *katso single nucleotide polymorphism*
- Solar** Kytöntäanalyysiohjelma, joka on suunniteltu määrällisiin ominaisuuksiin vaikuttavien geenien paikallistamiseen.
- SOM** (Self Organizing Map) *Itseorganisoiuva kartta* Yksi hermoverkkomenetelmien sovellus.
- somatic cell** *somaattinen solu* Mikä tahansa elion solu, joka ei kuulu iturataan.
- spliceosome** *spliseosomi* Proteiini/RNA-kompleksi, joka osallistuu intronien poistoon RNA:sta.
- splice site** *silmukoitumiskohta (leikkauskoh- ta)* RNA-sekvenssin kohta, jonka perusteella intronit tunnistetaan ja poistetaan lopullisesta lähettiRNA-sekvenssistä.
- splicing** *silmukoituminen (pujonta, liittäminen)* Prosessi, jossa DNA:n perusteella syntetisoidusta RNA:sta poistetaan intronialueet ja yhdistetään eksonialueet.
- sporadic case** *satunnainen tapaus* Ominaisuutta, yleensä sairautta, ilmentävä yksilö, jolla ei ole kyseistä ominaisuutta ilmentäviä esi-isiä.
- SRS** (Sequence Retrieval System) WWW-ympäristössä toimiva sekvenssihakutyökalu. Esimerkiksi: <http://srs.csc.fi:8002>
- SSCP** (single-strand conformational poly- morphism) katso single-strand conformational polymorphism.
- ssDNA** (single-stranded DNA) *yksijuosteinen (-nauhainen) DNA*
- ssRNA** (single-stranded RNA) *yksijuosteinen (-nauhainen) RNA*
- Staden** Molekyylibiologisen aineiston analyysityökalupaketti. Erityisen sovelias automaattisekvensaattoriaineiston käsitelyyn.
- sticky ended** Restriktioendonukleasien katkaisema DNA-kaksoisjuoste, jonka päässä on katkaisukohtaan epäsymmetrisyydestä johtuen lyhyt yksijuosteinen kohta. Tämä kohta voi liittyä toiseen DNA-molekyyliin, jossa on ensimmäiseen molekyyliin nähden vastinpäristä koostuva yksijuosteinen osuus. Katso flush ended.
- stochastic** *stokastinen, satunnainen*
- STR** (short tandem repeat) katso short tandem repeat.
- stringency** *ankaruus, vaativuus* Molekyylibiologiassa kuvaa usein sekvenssikoettimen sitoutumisolosuhteita tai sekvenssi-analyysissä vaadittavien pisteiden määrää esim. tarkasteluikkunassa.
- structure-activity relationship** (SAR) *rakenne-aktiivisuus-vuorovaikutus* SAR-menetelmissä käytetään tilastollisia menetelmiä rakenteellisten tekijöiden sekä ligandin tai proteiinin aktiivisuuden vastaavuuden arviointiin.
- STS** (sequene-tagged site) katso sequence-tagged site.
- subsequence** *alisekvenssi, osasekvenssi* Sekvenssi, joka on osa pidempää sekvenssiä. Esimerkiksi EST-sekvenssit tai proteiinidomeenisekvenssit.
- substitution matrix** *pisteytysmatriisi* Taulukko, jolla paikasta riippumattomissa sekvenssirinnastusmenetelmissä määritetään eri nukleiini- tai aminohappoyhdistelmien vaikutus rinnastuksen laatua tai todennäköisyyttä kuvaavaan pisteytykseen. Katso scoring matrix.

sum of pairs score *parisummapisteitys* Momen sekvenssin rinnastuksessa käytettävä pisteytysmenetelmä, jossa sarakkeen pistemäärä lasketaan kaikkien sarakkeissa olevien emäs/aminohappoparien parittaisten vertailujen pisteytysten summana. Esim. kolmen sekvenssin, A B C, tapauksessa parisummapisteitys on $(A+B)+(A+C)+(C+B)$.

supercoiling *superkierteisyys* DNA-kaksoiskierteen kiertyminen itsensä tai nukleosomin ympäri.

superfamily *superperhe* Proteiiniryhmä, jolla on yhteisiä piirteitä, esim. immunoglobuliinit.

supersecondary structure *supersekundäärirakenne* Säännöllinen, useammasta sekundäärirakenneyksiköstä koostuva rakenne.

SWISS-PROT Proteiinisekvenssitietokanta, jossa on pyritty hyvin jäseneltään proteiinisekvenssien ja niihin liittyvän tiedon kuvailuun. Sekvenssin lisäksi SWISS-PROT-tietueet sisältävät tietoa mm. proteiinin toiminnasta, alayksiköistä sekä vastaavista mutatoiduista proteiineista. <http://www.expasy.ch/sprot/>

T

target *kohde* Etsinnän kohteena oleva tuntematon sekvenssi tai molekyyli. Vrt. koetin (probe)

TATA-box *TATA-sekvenssi* Useista geeneistä löytyvä transkription aloituskohdan tunnistukseen liittyvä sekvenssikohta. TATA-jakson konsensussekvenssi on 5'-TATAAAA-3' ja se sijaitsee noin 25 nukleotidia ennen transkription aloituskohtaa.

taxon *taksoni* Mikä tahansa eliöluokittelun ryhmä.

TCP/IP (Transport Control Protocol/Internet Protocol) Yksi Internetin käyttämistä tiedonsiirtokäytännöistä.

telomere *telomeeri* Kromosomin pää.

template *malli, templaatti* Rakenne, joka

ohjaa toisen rakenteen syntyä. Esimerkiksi nukleotidisekvenssi, joka ohjaa vastin-nukleotideista muodostuvan nukleotidisekvenssin synteesiä. Molekyylihallitussa: tunnettu rakenne, jonka pohjalta toinen rakenne mallitetaan.

termination factor *lopetustekijä* Proteiini, joka vaikuttaa transkription päätekohdan tunnistamiseen RNA-synteesissä.

terminator codon *lopetuskodoni* Emäskolmikko, joka toimii RNA-ketjussa proteiinisynteesin lopetuskohdan merkkinä. Lope-tuskodoneita ovat UAA, UAG, UGA.

tertiary structure *tertiäärirakenne* Makromolekyyliin, esimerkiksi proteiinin, kolmiulotteinen laskostunut rakenne.

TIGR The Institute for Genomic Research. Ei-kaupallinen genomeihin ja geenituot-teisiin keskittynyt tutkimuslaitos. <http://www.tigr.org/>

topoisomerase *topoisomeraasi* Entsyymi, joka muuttaa DNA:n kolmiulotteista muotoa vaikuttamalla superkierteisyyteen (supercoiling) katkaisemalla DNA:n toisen tai molemmat juosteet.

topology *topologia* Muoto, muoto-oppi.

trait Perinnöllisyystieteessä: mikä tahansa geneettisesti määräytyvä ominaisuus.

transcription *transkriptio* DNA:n kopiaiminen proteiinisynteesin mallina toimivaksi lähetti-RNA:ksi. Katso splicing, cap.

transcriptome *transkriptomi* Solun tai eliön lähettiRNA-molekyylien kokonaisuus.

transgenic *siirtogeeninen, transgeeninen* Eliö, jonka perimään on siirretty toisen samaan tai eri lajiin kuuluvan yksilön perimää.

transition *transitio, siirtymä, muutos, virit-tyminen* DNA:n yhteydessä: mutaatio, jossa puriiniryhmän sisältävä nukleotidi korvautuu toisella puriiniryhmän sisältävällä nukleotidilla, tai vastaavasti pyrimidiiniryhmän sisältävä nukleotidi korvautuu toisella pyrimidiiniryhmän sisältävällä nukleotidilla. Katso transversio.

translation *translaatio, proteiinisynteesi*

Transkription jälkeinen tapahtuma, jossa lähettiRNA:n emäsjärjestys käännetään proteiinin aminohappojärjestykseksi.

transposon *transposoni* Pieni kopioitumis- ja liikkumiskykyinen DNA-molekyyli, joka liittyy satunnaiseen kohtaan genomissa.

transversion *transversio* Pistemutaatio, jossa puriini ryhmä korvautuu pyrimidiini ryhmällä tai päinvastoin. Katso transitiio.

trinucleotide repeat *kolmen nukleotidin toisto* DNA-sekvenssin toistoalue, joka saattaa kuulua myös koodaavaan alueeseen.

triple helix *kolmoiskierre* Esiintyy mm. kollageenirakenteissa. Myös DNA:n harvinainen muoto.

triplet, *kodoni, kolmikko* DNA:n nukleotidikolmikko, joka vastaa RNA:n kodonia.

trisomy *trisomia* Perinnöllinen häiriö, jossa yksilöllä on ylimääräinen kromosomi.

tRNA (transfer RNA) *siirtäjä-RNA* Pieni RNA-molekyyli, joka proteiinisynteesissä vastinkodoninsa ohjaamana tuo lähettiRNA:n osoittaman aminohapon ribosomilla tuotettavaan polypeptidiketjuun.

twilight zone *hämärän alue* Käytetään sekvenssivertailussa kuvaamaan sekvenssien samankaltaisuutta, kun niiden identtisyys on alle 25 prosenttia. Tämänasteinen samankaltaisuus voi olla satunnaista.

U

unassigned reading frame (URF) tehtävältään määrittämätön lukukehys.

unrooted tree *juureton puu* Puukaavio, jossa kaavion suunnalla ja osien keskinäisellä sijoittelulla ei ole merkitystä. Juureton puu kuvaa vain esittämänsä materiaalin välisiä suhteita.

untranslated region (UTR) *transloittumaton alue* cDNA:n ne osat, jotka eivät ilmenny proteiiniksi.

UPGMA (Unweighted Pair Group Method using arithmetic Averages). Yksinkertainen puukaavion muodostus algoritmi. Menetelmää on käytetty mm. geneettisiin etäisyyksiin perustuvasa fylogeniikka-analyysissä.

upstream *aikaisempi tai edeltävä sekvenssikohta* DNA-sekvenssin suunnan määrityksen vastainen suunta, 3'-5'.

URF (unassigned reading frame) katso unassigned reading frame.

UTR (untranslated region) katso untranslated region.

V

van der Waals interaction *van der Waals -vuorovaikutus* Atomien välinen vuorovaikutus, joka syntyy atomien toistensa varausjakaumaan aiheuttamista häiriöistä

van der Waals radius *van der Waals -säde* Etäisyys, jota lähempänä atomin ydintä van der Waals -vuorovaikutukset muuttuvat lähentyvän atomin suhteen hylkiviksi. Tarkoittaa käytännössä atomin sädettä.

variable number tandem repeat (VNTR) Polymorfista satelliitti-DNA:ta. Koostuu vaihtelevasta määrästä DNA-toistojaksoja.

Viterbi-algoritmi. *Viterbin algoritmi* Laskee todennäköisimmän polun kätkeyn Markovin mallin läpi. Viterbin algoritmia käytetään kätkeyten Markovin mallien yhteydessä parhaimman sekvenssirinnastuksen etsintään. Algoritmi muistuttaa sekä rakenteeltaan, että käytöltään paljolti dynaamisen optimoinnin (Needlemanin ja Wunschinin algoritmi) käyttöä pisteytysmatriisimenetelmien yhteydessä.

VNTR (variable number tandem repeat) katso variable number tandem repeat.

W

W2H GCG- ja EMBOSS-ohjelmistojen verkkoselainkäyttöliittymä. <http://www>.

w2h.dkfz-heidelberg.de/

Watson-Crick base pairs *Watsonin ja Crickin emäsparit* Watsonin ja Crickin DNA-mallissaan esittämä DNA:n emästen pariutumisjärjestys: adeniini-tyymiini, guaniini-sytosiini.

weight *painotus, paino* Bioinformatiikan algoritmeissa kerroin, jonka mukaan tiettyä tekijää painotetaan suhteessa muihin tekijöihin. Esimerkiksi profiilirinnastuksessa tiettyyn sekvenssikohtaan sidottu painoarvo.

wild type *villityyppi* Yleisin ilmiö tai genotyyppi. Alunperin: luonnosta löydetty yleisin muoto.

X

X-linked disease *X-kromosomiin liittyvä sairaus* Perinnöllinen sairaus, jonka aiheuttava geeni sijaitsee X-kromosomissa. X-kromosomi on toinen ihmisen sukupuolikromosomeista: normaalikaryotyypissä naisilla niitä on kaksi, miehellä yksi.

XML (Extensible Markup Language), Yleinen standardiformaatti, jolla voidaan kuvata rakenteista aineistoa ja dokumentteja tekstimuodossa. XML on suunniteltu erityisesti helpottamaan rakenteellisen aineiston siirtoa ja esittämistä WWW-järjestelmässä.

Xplor Ohjelma, jota käytetään biologisten makromolekyylien rakenteen hienonnutta ja määrittelylaskuihin röntgenkristallografiassa ja NMR-spektroskopiassa.

Y

YAC (yeast artificial chromosome) katso yeast artificial chromosome.

yeast artificial chromosome (YAC) *hiivan keinotekoinen kromosomi* Vektorijärjestelmä, jonka avulla voidaan kloonata pitkiäkin DNA-jaksoja hiivaan.

Z

Z-DNA *Z-DNA* DNA:n kolmas tun-

nettu rakennemuoto. Z-DNA:n kiertymissuunta on vastakkaisuuntainen (vasenkätinen kiertyminen) kuin A- ja B-DNA:lla (oikeakätinen kiertyminen). Toinen merkittävä ero A- ja B-DNA:han on juosteen fosfaattiryhmien mutkittava siksak-rakenne (tästä nimitys Z-DNA).

zinc finger (**metal-binding finger**)

sinkkisormi Joukko sekvenssimotiiveja, joille on yhteistä neljän histidiinin ja/tai kysteiniinivuketjun avulla tapahtuva sinkki-ionin sitominen. Sinkkisormet yleensä tunnistavat ja sitovat nukleiinihapoketjuja.

Z-score *Z-arvo* Yleinen sekvenssianalyysiohjelmien parametri, jota käytetään arvioitaessa tietokannasta löytyneen sekvenssin ja hakusekvenssin välistä suhdetta. Z-arvo kuvaa sitä todennäköisyyttä, jolla yhtä samankaltainen sekvenssi voidaan löytää sattumalta.

Z-value katso Z-score.

zygote *tsygootti* Kahden sukusolun yhdistymisessä syntyvä solu; hedelmöitynyt munasolu.

Å

Ångström Ångström (Å) = 0,1 nm. Molekyyliarakenteiden yhteydessä yleisesti käytetty pituuden mittayksikkö.

3 Nukleotidit ja nukleiinihapot

Numerot ja symbolit

(-)-strand (-)-juoste (DNA:ssa), ((-)-nauha, (-)-säie), *templaattijuoste* Syntyvä juoste on templaatille komplementaarinen.

(+)-strand (+)-juoste (DNA:ssa) ((+)-nauha, (+)-säie) Koodaava juoste, jonka kaltaisen syntyvä juoste tulee olemaan. Vastinjuoste, joka ei ole templaattijuoste. (+)-juosteella ja lähettiRNA:lla on sama sekvenssi (poikkeuksena T:n korvaus U:lla).

3'-end 3'-pää nukleiinihappojuosteessa Merkintä 3' viittaa sokeriryhmän hiiliatomien numerointiin. Katso kuvaa 2

3'-splice site 3'-silmutointikohta

5'-end 5'-pää nukleiinihappojuosteessa Merkintä 5' viittaa sokeriryhmän hiiliatomien numerointiin. Katso kuvaa 2

5'-splice site 5'-silmutointikohta

A

A-DNA A-DNA B-DNA:sta dehydraation kautta syntynyt rakennemuoto, jossa on 11 emäsparia/kierre. Katso B-DNA ja Z-DNA.

allele alleeli, *vastingeeni* Geenin vaihtoehtoiset muodot, jotka sijaitsevat samassa kohdassa kromosomissa. Alleeleja voi samalla diploidilla normaalikaryotyypillisellä yksilöllä olla kerrallaan vain kaksi, joko kaksi samanlaista (homotsygootti) tai kaksi erilaista (heterotsygootti).

alternative splicing *vaihtoehtoinen silmutointi, pujonta* Proteiinituotannon säätely lähettiRNA:n prosessoinnilla. Eksonien koodaamien lähettiRNA-sekvenssien liittäminen toisiinsa vaihtoehtoisista kohdista saa aikaan erilaisten proteiinien tuotannon, tai lähettiRNA:n, jota ei transloida.

Alu sequence *Alu-sekvenssi, yleinen toistojakso* Useissa ihmisen introneissa toistuva sekvenssi, joka tunnustetaan AluI-restriktiokohtien avulla. Homologiselle rekombinaatiolle alttiita kohtia.

amber mutation *amber-mutaatio* Lope-tuskodoniin (UAG) johtava mutaatio.

B

anticodon *antikodoni* Siirtäjä-RNA:n (tRNA:n) kolmen peräkkäisen kodonille vastakkaisen eli komplementaarisen nukleotidin muodostama tripletti, joka koodaa yhtä aminohappoa. Antikodoni tunnistaa kodonin, mikä mahdollistaa aminohappojen sijoittamisen oikealle kohdalle peptidiketjussa.

anticodon arm *antikodonisilmukka* tRNA:n antikodonin sisältävä osa.

antisense Geenille vastakkainen DNA-sekvenssi, joka toimii RNA:n mallina transkriptiossa.

autosome *autosomi* Muu kuin sukupuolikromosomi.

BAC (bacterial artificial chromosome) katso bacterial artificial chromosome.

bacterial artificial chromosome (BAC) *keinotekoinen bakteerikromosomi* Molekyylibiologiassa käytetty kantaja (vektori).

base emäs Nukleiinihappojen yhteydessä: Nukleiinihappojen perusyksikkö. DNA:n neljä emästyyppeä ovat adeniini (A), guaniini (G), tymiini (T) ja sytosiini (C). RNA:ssa tymiinin tilalla esiintyy urasiili (U).

base pair (bp) *emäspari* Nukleotidien emäkset, jotka kaksijuosteisessa DNA:ssa sitoutuvat toisiinsa vetysidoksilla. DNA:n neljä emästä ovat adeniini (A), guaniini (G), tymiini (T) ja sytosiini (C). Emäspareilla tarkoitetaan yleensä pareja A-T ja G-C, jotka ovat energeettisesti edullisimmat. RNA:ssa T korvautuu urasiililla (U).

B-DNA B-DNA DNA:n hydroitu rakennemuoto (Vertaa: A-DNA ja Z-DNA). Watsonin ja Crickin kuvaama DNA-rakennemuoto, jossa on 10,4 emäsparia/kierre.

blunt ended *tylppäpäinen* Kaksinauhaisen

DNA:n pää, jossa kumpikin nauha on yhtä pitkä.

C

CAAT-box *CAAT-sekvenssi* Useiden eukaryoottigeenien promoottorialueella sijaitseva konservoitunut DNA-sekvenssi, joka yleensä osallistuu transkription säätelyyn. CAAT-sekvenssi sijaitsee yleensä ns. TATA-laatikon 5'-puolella, n. 50 - 200 nukleotidia alavirtaan transkription aloituskohdasta.

CAI (codon adaptation index) katso codon adaptation index.

candidate gene DNA-alue, jonka sijainti ja sekvenssi saattavat liittyä tarkasteltavaan ominaisuuteen, esim. perinnölliseen sairauteen.

cap Kemiällinen rakenne, joka liitetään aitotumallisten esi-lähettiRNA:n 5'-päähän transkription jälkeen. Modifioitu guanosiinitrifosfaatti-7-metyyliguanosiini, joka on liittynyt 5'-5' trifosfaattisidoksella lähettiRNA:n 5'-päähän.

cap site lähettiRNA:n 5'-pää. Kohta, johon cap-rakenne liitetään, merkitään usein geeniin +1:llä.

cDNA (complementary DNA) *komplementaarinen DNA, vastakkainen DNA* DNA-juoste, joka on syntetisoitu käyttäen mallina RNA:ta.

cDNA library *cDNA-kirjasto* Kokoelma cDNA-molekyylejä, jotka on pakattu kantajaan (plasmidi, faagi).

centimorgan (cM) *senttimorgan* Geneettisen etäisyyden mitta (1/100 Morgania). 1 cM vastaa karkeasti ottaen 1 miljoonaa emäsparia DNA:ta ja yhden prosentin rekombinaatiofraktiota. Rekombinaatiofraktion ja karttaetäisyyden tarkka suhde määritellään karttafunktioiden avulla.

centromere *sentromeeri* Kromosomikäsivarsien kiinnittymiskohta, johon tuman jakautumisessa kiinnittyvät myös tuma-

sukkulan säikeet.

chiasma *kiasma* Meioosissa homologisten kromatidien välille muodostuva rakenne, joka mahdollistaa kromatidien geneettisen materiaalin vaihtuminen eli tekijäinvaihdunnan.

chromatid *kromatidi* Kahdentuneen kromosomin puolikas eli kromosomin puolikkaat ennen solunjakautumista edeltävää mitoosin metafaasivaihetta tai meioosin toisen jaon loppuvaiheita.

chromatin *kromatiini* Kromosomin pakkautunut DNA-rihma histoniproteiineineen. (Mikroskooppisesti: tiheästi pakkautunut heterokromatiini ja löysempi eukromatiini).

chromosomal map *kromosomikartta* Kromosomikuva, jossa on merkittynä geenimerkkien fyysinen sijainti.

chromosome *kromosomi* DNA-rihma, joka aitotumallisilla pakkautuu tiiviisti histoniproteiinien avulla. Pääosa solun geenistöstä on kromosomeissa.

cM (centimorgan) katso centimorgan.

coding strand koodaava juoste. Genomisen DNA:n juoste, jonka sekvenssi kopioituu lähettiRNA:han. Sekvenssiä julkaistaessa esitettävä emäsjärjestys.

codon *kodoni* lähettiRNA:ssa kolme peräkkäistä emästä, jotka proteiinisynteesissä vastaavat yhtä aminohappoa tai translaation lopetusta.

codon adaptation index (CAI) *kodonien adaptaatioindeksi* Synonyymikodonien käyttöön liittyvä mitta.

cohesive end *koheesiivinen pää* Kaksijuosteisen DNA:n pää, jossa toinen juoste on pidempi eli jatkuu yksijuosteisena.

complementary *komplementaarinen* Vastintai nukleinihappojuoste, jonka jokainen nukleotidimäs pariutuu (Watson-Crick -periaatteen mukaisesti) toisen juosteen kanssa.

complementary base *vastinemäs* Kaksijuosteisessa DNA:ssa jokaista A:ta vastaa T ja jokaista G:tä C.

contig *jatkumo* Lyhyistä, osittain päällekkäisistä sekvensseistä yhdistelemällä muodostettu yhtenäinen sekvenssi.

cosmid *kosmidi* Kantaja (vektori), joka on tehty yhdistämällä λ -faagin DNA:ta bakteerin plasmidiin. Käytetään usein pitkien DNA-fragmenttien kloonauksessa.

CpG island *CpG-saareke* Runsaasti C- ja G-nukleotideja sisältävä alue. Usein lähellä aktiivisesti ilmenettäviä geenejä.

crossing over *tekijäinvaihdunta* Geneettisen materiaalin vaihtuminen vastinkromosomien välillä kromatidirihojen ristiinmenon yhteydessä meiosisin 1. jaossa.

D

deletion *poistuma, deleetio, puuttuma* Nukleiinihappojuosteiden osan ja siinä olevien geenien poistuminen. Mutaatiotyyppi.

depurination *depurinaatio* Puriinin poisto nukleiinihappoista.

DGGE technique DNA-geelelektroforeesi denaturoivassa pH-gradientissa. Käytetään geenien monimuotoisuuden analysoinnissa.

dideoxynucleotide *dideoksinukleotidi* Nukleotidi, jossa deoksiriboosi on korvattu dideoksiriboosilla. DNA-polymeraasit eivät pysty liittämään uusia nukleotideja dideoksinukleotideihin, mistä johtuen niitä voidaan käyttää polymerisaation pysäyttämiseen.

dideoxynucleotide sequencing *dideoksinukleotidisekvensointi* Sangerin entsymaattinen DNA:n sekvensointimenetelmä, jossa sekvensoitavalle juosteelle syntetisoidaan dideoksinukleotidien läsnäollessa leimattuja, eri pituisia juosteita. Pituusjärjestyksen avulla voidaan määrittää DNA:n emäsjärjestys.

differential display (DD) lähettiRNA-populaatioista RT-PCR:llä muodostettu kokoelma eri pituisia DNA-fragmentteja, jotka sekvensointigeelissä eroteltuna osoittavat samankaltaisesti tai eri tavoin

ilmeneviä geenejä.

direct repeat *suora toisto* Samassa suunnassa toistuva emäsjärjestys (esim. ACT ACT ACT). Katso palindromi.

DNA (deoxyribonucleic acid) *deoksiribonukleiinihappo*

DNA microarray *DNA-siru* Alustalle (lasi tai kalvo) kiinnitetty kokoelma DNA-molekyylejä, jotka voidaan hybridisoida liuoksessa olevan nukleiinihappokokoelman kanssa.

domain *domeeni* Proteiinin rakenteen itsenäisesti poimuttava yksikkö.

donor *luovuttaja, donori* Atomi, molekyyli tai solu, joka luovuttaa vastaanottajalle esimerkiksi elektronin, molekyylin tai molekyyliyhymän

double helix (Watson-Crick model)

kaksoiskierre, Watsonin ja Crickin malli Molekyylibiologiassa: DNA:n vastinjuosteet kiinnittyvät emäsosistaan toisiinsa kaksoiskierteiseksi molekyyliarakenteeksi.

double-stranded *kaksijuosteinen* (-nauhainen, -säikeinen) Nukleiinihapporakenne, joka muodostuu kahdesta emäspariutuneesta vastinjuosteesta.

downstream *alavirran puoleinen, 5'-3' -suuntaan etenevä, 3'-puolella sijaitseva*; proteiineissa karboksiterminalipään puoleinen.

dsDNA (double-stranded DNA) *kaksijuosteinen* (-nauhainen, -säikeinen) DNA

dsRNA (double-stranded RNA) *kaksijuosteinen* (-nauhainen, -säikeinen) RNA

duplication (*kromosomin osan kahdentuminen*) Kromosomin osan kahdentuminen. Mutaatiotyyppi.

E

EMBL Euroopan molekyylibiologian laboratorion ylläpitämä DNA-sekvenssietokanta, jota ristiinpäivitetään DDBJ:n ja GenBankin kanssa. <http://www.ebi.ac.uk/embl/index.html>

enhancer *tehostaja, voimistaja* Molekyylibiologiassa: transkriptiotekijä, joka yleensä aktivoi transkriptiota.

EST (expressed sequence tag) katso expressed sequence tag.

expressed sequence tag (EST) *ilmenevän geenin osa* EST-sekvenssi on solun tiettyssä tilassa ilmenevän geenin satunnaisesti ja usein vain osittain sekvensoitu cDNA-klooni.

F

flanking sequence kohdesekvenssin viereiset alueet (5'-flank, 3'-flank).

flush ended *kohesiivinen pää* Kaksijuosteinen DNA, jossa toinen juosteista on pidempi. Katso sticky ended.

frameshift Lukukehyksen muuttuminen esim. poistuman tai lisäyksen seurauksena.

G

genomic DNA *genominen DNA* Solun DNA-fraktio, joka muodostuu kromosomeista (ei siis sisällä soluelinten DNA:ta).

H

heteroduplex *heterodupleksi* DNA-molekyyli, jossa juosteet ovat erilaisia. Juosteilla voi olla eri alkuperä, jolloin komplementaariset juosteet on saatu yhdistymään in vitro tai dupleksi muodostuu lähettiRNA:sta sekä vastaavasta DNA:sta.

heteroduplex analysis (mismatch analysis) *heterodupleksianalyysi* Eri alkuperää olevien DNA-juosteiden annetaan liittyä toisiinsa heterodupleksiksi, minkä jälkeen erot kartoitetaan. Kartoitusmenetelmiä on useita.

high throughput genome (HTG) Viimeistelemtöntä sekvenssiaineistoa, josta kuitenkin on hyötyä esimerkiksi sekvenssihauissa. Tietokannassa vähintään kahden kiloemäksen palasina. Nopeasti tuotettua raaka-aineistoa, jota ei ole

varmennettu.

hnRNA (heterogeneous nuclear RNA) *heterogeeninen tumen RNA* Lyhytikäinen ja suurimolekyylinen lähettiRNA:n esiaste.

homeobox *homeolaatikko* Konservoitunut DNA-jakso, joka alunperin löytyi banaanikärpäsen useista homeoottisista tai segmentaatiomutaatioita aiheuttavista geneistä.

hormone response element (HRE) *hormonivastealue* DNA-sekvenssi, jonka hormonireseptori tunnistaa ja jonka ilmenemistä reseptori säätelee.

hot spot DNA-alue, jossa muita alueita selvästi useammin tapahtuu evolutiivisia muutoksia, esim. mutaatioita.

HRE (hormone response element) katso hormone response element.

HTG (high throughput genome) katso high throughput genome.

hybridization *yhdistyminen, hybridisaatio*
1. Tapahtuma, jossa kaksi vastakkaista nukleiinihappujuostetta liittyy yhteen. Käytetään geenitekniikassa nukleiinihappojen tunnistusmenetelmänä. Voidaan käyttää diagnostiikassa tiettyjen DNA-jaksojen osoittamiseen näytteestä (katso: probe, koetin). 2. Risteytyminen. 3. Solujen yhdistäminen keinotekoisesti.

I

indel (**insertion or deletion**) lisäys tai poistuma DNA- tai proteiinimolekyylistä.

in frame *lukukehyksessä, lukukehyksen säilyttävä*

initiation codon *aloituskodoni* lähettiRNA:n emäskolmikko (yleensä 5'-AUG), josta proteiinisynteesi alkaa.

initiator tRNA *aloitus-tRNA* Aminohapposynteesin aloittava tRNA.

intron *introni* Geeninsisäinen exonien välissä oleva ei-ilmentyvä DNA-alue. Introni kopioidaan RNA:ksi, mutta poistetaan siitä ennen proteiinisynteesiä. Katso:

exon ja splicing.

inverse PCR *käänteinen PCR*

inverted repeat (IR)*käänteistoisto* Nukleotidisekvenssi, joka on käänteinen verrattavalle sekvenssille. Useimmiten alle 50 emästä. IS-elementtien osia. (IS-elementti, insertion sequence, on transposonityyppi).

IR (inverted repeat) katso inverted repeat.

K

kb (kilobase) katso kilobase (kb), kilobasepair (kbp).

kbp (kilobasepair) katso kilobase (kb), kilobasepair (kbp).

kilobase, kilobasepair (kb, kbp)*kiloemäs, kiloemäspari* Tuhannen nukleotidin pituinen DNA- tai RNA-jakso.

Kozak sequence Proteiinisynteesin lähettiRNA-aloituskodonin koodaaman aminohapon ympäriltä valmiista proteiineista löytyvä konservoitunut jakso.

L

lagging strand *viivästynyt ketju, jälkijuoste* DNA:n kahdentuessa toinen syntyvistä juosteista. Syntyy pieninä, myöhemmin yhteenliitettävänä palasina.

LCR (locus control region)katso locus control region.

leading strand *johtava ketju, johtajuoste* DNA:n kahdentuessa uusista juosteista se, joka syntyy jatkuvana kokonaisuutena.

LINE (long interspersed repeat element) katso long interspersed repeat element.

locus control region (LCR) Alue, jolla geenin säätelyä ohjaavat elementit (promootorit ja tehostajat) sijaitsevat.

lod score Kymmenkantainen logaritmi kytkentää testaavasta uskottavuusosamäärästä. Kahden lokuksen kytkentää esittävä tilastotieteellinen arvo. Jos arvo on kolme tai suurempi, katsotaan lokusten olevan kytkettyneitä toisiinsa.

long interspersed repeat element (LINE)

Usean kiloemäksen pituinen nisäkkäiden genominen toistojakso.

long terminal repeat (LTR) Usean sadan emäksen pituiset käänteiset toistojaksot transposonien ja virusten esiasteiden DNA-ketjun päissä. Vaikuttavat transposonien ja viruksien esiasteiden liittymiseen isäntä-DNA:han.

LTR (long terminal repeat) katso long terminal repeat.

M

marker merkkijakso Genominen tai soluelimen DNA:n emäsjärjestys, joka eroaa tarpeeksi yksilöiden välillä, jotta sen periytyminen sukulinjassa ja/tai erilaisissa soluissa voidaan jäljittää. Merkkijakso voi olla geenissä tai geenien ulkopuolella.

Megabase (Mb)Miljoona emästä. Nukleotidisekvenssin pituuden mittayksikkö.

Megabasepair (Mbp) *megaemäs, megaemäspari* Miljoona emästä pitkä nukleiinihappojakso.

microsatellite mikrosatelliitti Genomissa runsaita yhdestä viiteen emäksen pituisia toistojaksoja, joiden lukumäärä vaihtelee yksilöittäin. Voidaan käyttää merkkijaksosoina.

minisatellite minisatelliitti Noin 10-100 emästä pitkiä toistojaksoja, joiden määrä vaihtelee yksilöittäin. Yleensä lähellä kromosomin päätä. Voidaan käyttää merkkijaksosoina.

minus-strand miinus-juosteinen (-nauhainen, *säikeinen*) RNA-virusten komplementaarinen DNA-juoste. Katso (-)-strand.

mismatch repair yhteensopimattomuuden korjaaminen DNA-synteesin aikana väärin pariutuneiden nukleotidien etsiminen ja korvaaminen.

missense mutation Mutaatio, joka muuttaa kodonin nukleotidijärjestystä niin, että sen koodaama aminohappo muuttuu.

monocistronic mRNA lähettiRNA, joka translaatiossa tuottaa vain yhden

polypeptidiketjun.

mRNA (messenger RNA) *lähetti-RNA* Geenin DNA-juosteelle komplementaarinen yksijuosteinen RNA-molekyyli, joka syntyy transkriptiossa. LähettiRNA ohjaa proteiinisynteesiä ja määrää polypeptidin aminohappojärjestyksen.

N

non-coding strand *koodaamaton juoste* Geenille komplementaarinen DNA-juoste.

nucleic acid *nukleinihappo* DNA ja RNA ovat nukleotideista rakentuvia nukleinihappoja.

nucleoside *nukleosidi* Nukleinihapon ja nukleotidin osa. Puriini- tai pyrimidiiniemäs + pentoosisokeri.

nucleosome *nukleosomi* Kromatiinin kromosomiksijärjestymisen perusyksikkö. Koostuu n. 200 emäsparista ja histoniproteiiniytimestä.

nucleotide *nukleotidi* Nukleosidi + fosfaatti. Nukleinihapon rakenneyksikkö.

O

oligonucleotide *oligonukleotidi* Korkeintaan muutamia kymmeniä emäksiä sisältävä DNA-sekvenssi.

open reading frame (ORF) *avoin lukukehys* Mahdollisesti proteiiniksi käännettävä DNA-sekvenssi, joka alkaa aloituskodonilla ja päättyy lopetuskodoniin. Katso: lukukehys, reading frame

operator *operaattori* DNA-alue, johon sitoutumalla estäjäproteiini estää transkription aloittamisen lähellä sijaitsevasta promootorista.

operon *operoni* Bakteerien geenitoiminnan yksikkö, johon kuuluu promootori, operaattori, sekä yksi tai useampia rakennegeenejä.

ORF (open reading frame) katso open reading frame.

orthology *ortologia, ortologisuus* Lajiutumisen seurauksena syntynyt

samankaltaisuus.

overlapping genes *päällekkäiset geenit* Erilisiä geenejä, joiden ilmennettävät alueet ovat osittain päällekkäiset ja joita luetaan eri lukukehyksessä.

P

palindrome *palindromi, käänteinen toistojakso* DNA-jakso, joka on sama luettuna kaksoskierteen kummastakin juosteesta. Esimerkiksi esitumallisissa lähellä transkription lopetuskohtaa sekä useimpien restriktioentsyymien tunnustussekvensseissä.

PCR (polymerase chain reaction) katso polymerase chain reaction.

polymerase chain reaction (PCR) *polymerasiketjureaktio* Tehokas tapa monistaa DNA:ta in vitro. Perustuu lämpövaihteluita kestävään DNA-polymeraasiin.

polymorphism *monimuotoisuus, vaihtelu, polymorfismi, polymorfia* Populaatiossa on kaksi tai useampia ilmiäsuja (saman geenin suhteen), joita ei voi selittää pelkästään mutaatioilla.

post-transcriptional modification *transkription jälkeinen muokkaaminen* Transkriptiossa syntyneen RNA:n rakenteeseen ennen translaatiota tehtävät muutokset. Esimerkiksi intronialueiden poisto, silmukointi (pujonta).

Pribnow box *Pribnow-sekvenssi* DNA-sekvenssijakso (TATAAT), joka ilmaisee bakteerien DNA:ssa RNA-synteesin aloituskohdan. Sijaitsee 5'-päässä, 10 emästä ennen synteisin aloituskohtaa. Voi esiintyä muuntuneena.

primer *aluke* Lyhyt polynukleotidiketju, jota käytetään käynnistämään DNA:n kahdentuminen.

promoter *promootori* DNA-sekvenssin kohta, johon RNA-polymeraasi sitoutuu ennen transkription aloitusta. Esimerkiksi Pribnow- ja TATA-sekvenssit.

pseudogene *pseudogeeni, valegeeni* Ilmestymätön DNA-jakso, jonka emäsjärjestys

muistuttaa jonkin geenin emäsjärjestystä.

purine puriini Typpeä sisältävä kaksoisringsrakenteinen orgaaninen yhdiste. Toinen nukleinihappojen emäsryhmistä. Puriinin sisältäviä nukleotideja ovat adeniini ja guaniini.

pyrimidine pyrimidiini Typpeä ja aromaattisen rengasrakenteen sisältävä orgaaninen yhdiste. Toinen nukleinihappojen emäsryhmistä. Pyrimidiinin sisältäviä nukleotideja ovat sytosiini, tymiini sekä urasiili.

R

RACE (rapid amplification of cDNA ends) katso rapid amplification of cDNA ends.

rapid amplification of cDNA ends

(RACE) Tekniikka, jolla cDNA:n 5'- tai 3'-pää saadaan PCR-monistettua, kun osa päiden välisestä sekvenssistä on selvillä.

reading frame lukukehys Kodonien rajakohtien ja lukusuunnan määrittely DNA-sekvenssissä. Koska DNA-sekvenssi voidaan jakaa kolmen emäksen mittaiseksi kodonisarjaksi kolmella tavalla kahteen eri suuntaan, on jokaisella DNA-sekvenssillä 6 mahdollista lukukehystä, joista yksi on translaatioissa käytetty lukukehys.

regulatory element säätelyosa Geenin ilmentymistä säätelevä (vähentävä tai voimistava) DNA-sekvenssin osa.

replication replikaatio, kopioituminen Biologiassa: solun jakautumista edeltävä perimäaineksen (DNA:n) kahdentuminen.

replicon replikoni, replikaatioyksikkö Perimän osa, joka pystyy kopioitumaan itsenäisesti. Nukleinihappoketjun osa, joka kopioituu saman replikaation aloituskohdan ohjaamana. Esitumallisilla on kromosomissa yksi replikoni, aiotumallisilla useampia.

restriction map restriktiokartta Kaavio, joka osoittaa restriktioentsyymien tunnistuskohtien sijainnin DNA:ssa.

reverse open reading frame (RORF) kään-

teinen avoin lukukehys.

RORF (reverse open reading frame) katso reverse open reading frame.

rRNA (ribosomal RNA) ribosomaalinen

RNA Ribosomien rakenneosia. Suurin osa solun RNA:sta on erikokoisia ribosomaalisia RNA-molekyylejä.

RT-PCR (reverse transcriptase polymerase chain reaction) *käänteistranskripti* *polymeraasiketjureaktio* Käänteiskopioijaentsyymiä käyttävä PCR.

S

satellite DNA satelliitti-DNA DNA-sekvenssin kohta, joka poikkeaa huomattavasti ympäristöstään yleensä tiheästi toistuvan sekvenssijakson johdosta ja jonka tehtävää ei tunneta.

self-splicing Itsesilmukointi Silmukoiminen itsenäisesti ilman ulkopuolisten molekyylien apua.

sense-strand kaksinauhaisen DNA:n koodaava juoste. Tässä juosteessa nukleotidijärjestys on sama kuin lähettiRNA:ssa.

sequence-tagged site (STS) *sekvenssimerkitty alue* Muutaman sadan emäsparin pituinen DNA-sekvenssijakso, joka esiintyy genomissa vain yhdessä tunnetussa kohdassa ja on tunnistettavissa PCR-tekniikalla. Sekvenssitunnusalueita käytetään tunnistettaessa ja yhdisteltäessä eri lähteistä saatua sekvenssiainestoa.

Shine-Dalgarno sequence Shine-Dalgarno -sekvenssi Bakteri-lähettiRNA:ssa yleinen sekvenssijakso, jonka perusmuoto on 5'-AGGAGG-3'. Se tai samankaltainen sekvenssi esiintyy yleensä noin 7 emästä ennen aloituskodonia.

short interspersed repeat element (SINE) Lyhyt genominen toistojakso.

short tandem repeat (STR) lyhyt peräkkäinen toistojakso.

shotgun sequencing satunnaissekvensointi (haulikkosekvensointi) Menetelmä, jossa genomi pilkotaan satunnaisen pituisiksi paloiksi, joiden (kaikkien tai osan)

emäsjärjestys määritetään.

silent mutation *hiljainen mutaatio, peittyvä mutaatio* DNA-muutos, joka ei näy geenin koodaaman proteiinin sekvenssissä.

SINE (short interspersed repeat element) katso short interspersed repeat element.

single-gene disorder Tauti tai tila, jonka aiheuttaa yksi geeni.

Single Nucleotide Polymorphism (SNP) *yhden emäksen monimuotoisuus* Lajin sisäinen nukleotidisekvenssin tietyssä kohdassa esiintyvä vaihtelu.

single-strand conformational polymorphism (SSCP) Pistemutaatioiden havainnointitekniikka. Tietty sekvenssialue PCR-monistetaan ja ajetaan geelissä. Mutaatio voi muuttaa rakennetta, mikä muuttaa molekyylin liikkuvuutta geelissä.

site-directed mutagenesis *kohdennettu mutageneesi* Yleisnimitys laboratorion menetelmille, joilla DNA-sekvenssiin tehdään haluttuun kohtaan yksittäisen tai useamman emäksen muutos.

site-specific mutagenesis katso site-directed mutagenesis.

SNP (single nucleotide polymorphism) *katso single nucleotide polymorphism*

splice site *silmukoitumiskohta (leikkauskoh- ta)* RNA-sekvenssin kohta, jonka perusteella intronit tunnistetaan ja poistetaan lopullisesta lähettiRNA-sekvenssistä.

splicing *silmukoituminen (pujonta, liittäminen)* Prosessi, jossa DNA:n perusteella syntetisoidusta RNA:sta poistetaan intronialueet ja yhdistetään eksonialueet.

SSCP (single-strand conformational polymorphism) katso single-strand conformational polymorphism.

ssDNA (single-stranded DNA) *yksijusteinen (-nauhainen) DNA*

ssRNA (single-stranded RNA) *yksijusteinen (-nauhainen) RNA*

sticky ended Restriktioendonukleasien katkaisema DNA-kaksoisjuoste, jonka päässä on katkaisukohdan epäsym-

metrisyydestä johtuen lyhyt yksijusteinen kohta. Tämä kohta voi liittyä toiseen DNA-molekyylisiin, jossa on ensimmäiseen molekyylisiin nähden vastinpareista koostuva yksijusteinen osuus. Katso flush ended.

STR (short tandem repeat) katso short tandem repeat.

STS (sequene-tagged site) katso sequence-tagged site.

supercoiling *superkierteisyys* DNA-kaksoiskiirteen kiertyminen itsensä tai nukleosomin ympäri.

superfamily *superperhe* Proteiiniryhmä, jolla on yhteisiä piirteitä, esim. immunoglobuliinit.

T

TATA-box *TATA-sekvenssi* Useista geeneistä löytyvä transkription aloituskohdan tunnistukseen liittyvä sekvenssikohta. TATA-jakson konsensussekvenssi on 5'- TATAAAA-3' ja se sijaitsee noin 25 nukleotidia ennen transkription aloituskohtaa.

telomere *telomeeri* Kromosomin pää.

terminator codon *lopetuskodoni* Emäskolmikko, joka toimii RNA-ketjussa proteiinisynteesin lopetuskohdan merkkinä. Lopetuskodeineita ovat UAA, UAG, UGA.

transcription *transkriptio* DNA:n kopiaaminen proteiinisynteesin mallina toimivaksi lähetti-RNA:ksi. Katso splicing, cap.

transcriptome *transkriptomi* Solun tai elion lähettiRNA-molekyylien kokonaisuus.

transition *transitio, siirtymä, muutos, virityminen* DNA:n yhteydessä: mutaatio, jossa puriiniiniryhmän sisältävä nukleotidi korvautuu toisella puriiniiniryhmän sisältävällä nukleotidilla, tai vastaavasti pyrimidiiniryhmän sisältävä nukleotidi korvautuu toisella pyrimidiiniryhmän sisältävällä nukleotidilla. Katso transversio.

transposon *transposoni* Pieni kopioitumista ja liikkumiskykyinen DNA-molekyyli, joka liittyy satunnaiseen kohtaan genomis-

sa.

transversion *transversio* Pistemutaatio, jossa puriiniiryhmä korvautuu pyrimidiiniryhmällä tai päinvastoin. Katso transiatio.

trinucleotide repeat *kolmen nukleotidin toisto* DNA-sekvenssin toistoalue, joka saattaa kuulua myös koodaavaan alueeseen.

triplet, *kodoni, kolmikko* DNA:n nukleotidikolmikko, joka vastaa RNA:n kodonia.

U

unassigned reading frame (URF) tehtävältään määrittämätön lukukehys.

untranslated region (UTR) *transloitumaton alue* cDNA:n ne osat, jotka eivät ilmenny proteiiniksi.

upstream *aikaisempi tai edeltävä sekvenssikohta* DNA-sekvenssin suunnan määrittämisen vastainen suunta, 3' - 5'.

URF (unassigned reading frame) katso unassigned reading frame.

UTR (untranslated region) katso untranslated region.

V

variable number tandem repeat (VNTR) Polymorfista satelliitti-DNA:ta. Koostuu vaihtelevasta määrästä DNA-toistojaksoja.

VNTR (variable number tandem repeat) katso variable number tandem repeat.

W

Watson-Crick base pairs *Watsonin ja Crickin emäsparit* Watsonin ja Crickin DNA-mallissaan esittämä DNA:n emästen pariutumisjärjestys: adeniini-tymiini, guaniini-sytosiini.

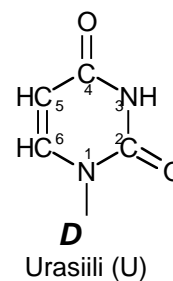
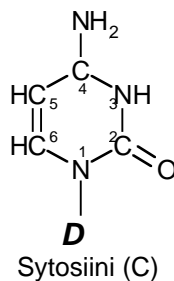
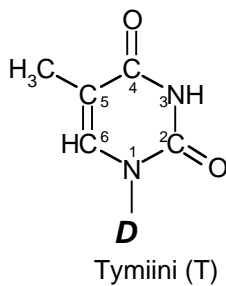
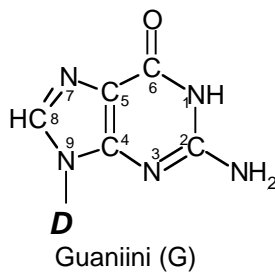
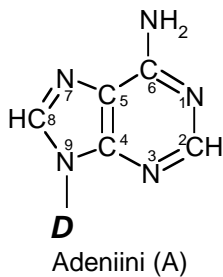
Z

Z-DNA *Z-DNA* DNA:n kolmas tunnettu rakennemuoto. Z-DNA:n kiertymissuunta on vastakkaisuuntainen

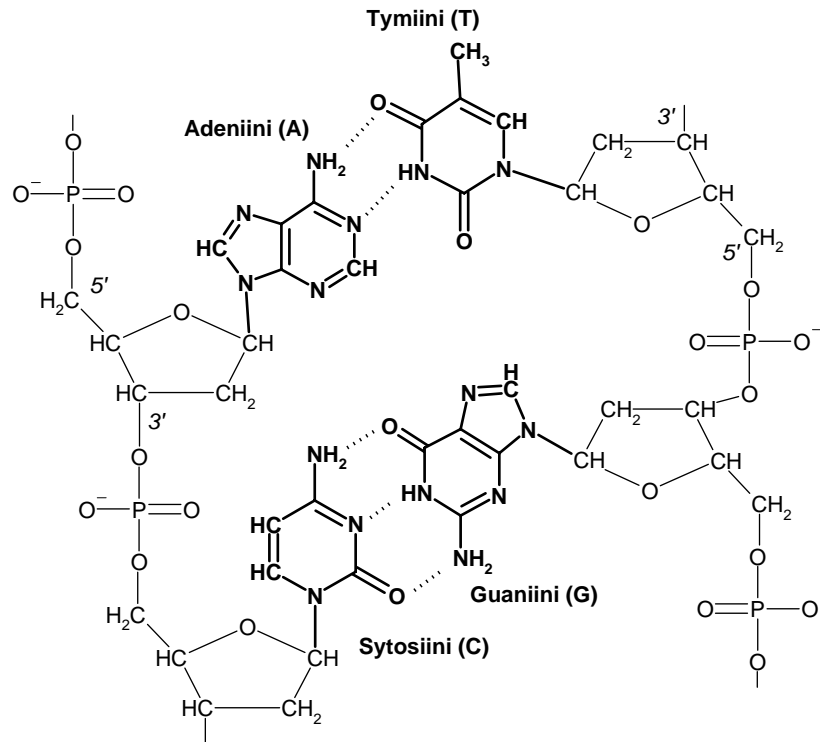
(vasenkätinen kiertyminen) kuin A- ja B-DNA:lla (oikeakätinen kiertyminen). Toinen merkittävä ero A- ja B-DNA:han on juosteen fosfaattiryhmien mutkittava siksak-rakenne (tästä nimitys Z-DNA).

4-thiouridine			4-tiouridiini
5-methylcytosine		5mC	5-metyylisytyosiini
6-dimethyladenine			6-dimetyyliadeniini
adenine	Ade	A	adeniini
cytosine	Cyt	C	sytyosiini
dihydrouracil		<i>UH₂</i>	dihydrourasiili
guanine	Gua	G	guaniini
hypoxanthine			hypoksantiini
inosine		I	inosiini
methylguanine		mG	metyyliguaniini
methylinosine		mI	metyyli-inosiini
N,N-dimethylguanine		<i>m₂G</i>	N,N-dimetyyliguaniini
N ⁶ -isopentenyladenine			N ⁶ -isopentenyliadeniini
pseudouracil		<i>ψ</i>	pseudourasiili
thymine	Thy	T	tymiini
uracil	Ura	U	urasiili
xanthine			ksantiini

Taulukko 1: Nukleiinihappoihin liittyviä kemiallisia yhdisteitä ja niiden lyhtenteitä



Kuva 1: Nukleiinihappoissa esiintyvät emäkset. Rengasrakenteiden atomit on numeroitu systemaattisesti. D osoittaa paikan, johon sokeriryhmä liittyy



Kuva 2: DNA:n kemiallinen rakenne. DNA:n nukleinihappojen neljä eri emästyppiä, adeniini, tymiini, sytosiini ja guaniini, on esitetty rakennekuvassa lihavoituina. Emäsparien (A-T ja G-C) väliset vetysidokset on esitetty katkoviivoilla.

4 Proteiinit

A

α -carbon *α -hiili* Katso selitystä aminohappojen osien nimeämisestä. Katso kuvaa 4

α -helix *α -kierre* Proteiinin sekundaäri-rakenne, jossa proteiinipääketju muodostaa kierteen. Kierrettä pitävät koossa ensimmäisen (n) peptidisidoksen karbonyyliryhmän (C=O) ja neljännen (n+4) peptidisidoksen amidiryhmän (H-N) väliset vetysidokset.

amphipathic helix *amfipaattinen kierre* Proteiinarakenteessa esiintyvä kiertekerä rakenne, jossa vesihakuiset ja vesipakoiset aminohapot sijoittuvat kierteen vastakkaisille puolille.

B

β -hairpin *hiusneularakenne* Proteiinin sekundaäri-rakenne, joka koostuu kahdesta vastakkaisuuntaisesta β -säikeestä ja niitä yhdistävästä silmukasta.

β -sheet *β -levy, β -taso* Vierekkäisten β -säikeiden muodostama vetysidosten stabiiloima proteiinin sekundaäri-rakenne. Taso on yhdensuuntainen tai vastakkaisuuntainen sen mukaan, kulkevatko vierekkäisten säikeiden peptidiketjut saman- vai erisuuntaisesti.

β -strand *β -säie* Proteiinin sekundaäri-rakenne, jossa pääketju on lähes maksimaalisesti oienneena. β -säikeet voivat esiintyä yksittäin, mutta yleensä ne sijoittuvat proteiinarakenteissa vierekkäin muiden β -säikeiden kanssa muodostaen β -levyjä.

β -turn *β -käännös* Proteiinin sekundaäri-rakenne, jossa peptidiketju kääntyy jyrkästi kolmen aminohapon matkalla. Vetysidos pitää käännöstä koossa. β -käännöksiä tunnetaan useita eri tyyppiä.

C

CATH (Class, Architecture, Topology and Homologous superfamily). Tietokanta, jossa PDB-rakenteiden proteiinidomeenit

on luokiteltu rakenteellisten ominaisuuksien mukaan. <http://www.biochem.ucl.ac.uk/bsm/cath/>

chaperone *kaperoni* Proteiini, joka avustaa toisen proteiinin laskostumista.

coiled coil *kiertynyt kierre* Proteiini-rakenne, jossa samansuuntaiset α -kierteet kiertyvät toistensa ympäri.

D

disulphide bond *disulfididisidos, rikkisilta* Kahden rikkiatomin välinen kovalenttinen sidos. Proteiineissa kysteiniinivuketjujen välisillä disulfididisidoksilla eli rikkisilloilla on usein merkittävä vaikutus proteiinin kolmiulotteiseen rakenteeseen. Katso disulphide bridge.

disulphide bridge *rikkisilta* Kahden kysteiniinivuketjun rikkiatomien välinen disulfididisidos. Proteiinarakenteissa rikkisillat voivat sitoa kovalenttisesti yhteen kaksi sekvenssin eri kohdissa olevaa kysteiniä ja siten vaikuttaa proteiinin rakenteeseen ja sen vakauteen.

G

Greek key motif *Greek key -motiivi* β -säikeiden muodostama rakenteellinen motiivi.

H

helix-loop-helix (HLH) motif *kierre-silmukka-kierre -motiivi* Kaksi α -kierrettä, joita erottaa silmukka. Transkriptiotekijöihin liittyvä motiivi, joka tunnistaa spesifisiä DNA-sekvenssejä ja sitoutuu niihin.

helix-turn-helix (HTH) motif *kierre-käännös-kierre -motiivi* Kaksi amfipaattista α -kierrettä, joita erottaa lyhyt β -käännös. Transkriptiotekijöihin liittyvä motiivi, joka tunnistaa spesifisiä DNA-sekvenssejä ja sitoutuu niihin.

histone *histoni* DNA:han sitoutuva, sitä stabiiloiva ja sen toimintaa säätelevä pro-

teini. Histonit jaotellaan viiteen eri tyyppiin: H1, H2a, H2b, H3 ja H4. Histonien avulla aitotumallisten (eukaryootien) DNA pakkautuu tiiviisti.

homeodomain *homeodomeeni* Homeolaatikko-sekvenssin koodaama proteiinin osa.

hydrophilic residue *vesihakuinen aminohappo*

hydrophobic zipper (leucin zipper)

hydrofobinen vetoketju DNA:han sitoutuvissa proteiineissä esiintyvä rakenne, jolla proteiini tunnistaa sitoutumiskohdan DNA:ssa. Rakenne muodostuu kahdesta rinnakkaisesta kierteestä, joissa molemmissa joka seitsemäs aminopetto on vesipakoinen (yleensä leusiini).

I

isoschizomer *isoskitsomeeri* Restriktioentsyymit, jotka tunnistavat saman sekvenssin, ovat keskenään isoskitsomeereja.

N

nuclear localization signal *tumaanohjaussignaali* Aitotumallisissa tumaan kuljetettävien proteiinien signaalina toimiva peptidisekvenssi.

nucleotide-binding domain nukleotidin sitova domeeni proteiinissa.

P

peptide *peptidi* Lyhyt aminohappoketju, pieni proteiini tai proteiinin pilkottu osa.

peptide bond *peptidisidos* Kovalenttinen sidos, jolla aminohapon pääketjun karboksyyli-ryhmä liittyy toisen aminohapon aminoryhmään.

pleated sheet *aaltoileva tai poimuinen levy* β -tason tyypillinen rakenne, jossa proteiinin pääketjun muoto aaltoilee. Yksittäisen β -nauhan perättäiset α -hiilet sijaitsevat vuorotellen β -tason keskikohdan eri puolilla ja muodostavat β -tasolle tyypillisen aaltoilevan hienorakenteen.

post-translational modification *translaation*

jälkeinen muokkaaminen Translaatiossa tuotettuun aminohappoketjuun tehtävät muutokset, esim. signaalisekvenssien poistaminen tai metyyli-, fosfaatti- ja sokeriryhmien liittäminen kovalenttisesti proteiiniin.

proteome *proteomi* Eliön tai solun tuottamien proteiinien kokonaisuus.

Q

quaternary structure *kvaternäärirakenne*

Useammasta eri aminohappoketjusta koostuvan proteiinikompleksin kokonaisrakenne. Proteiinikompleksin alayksikköjen sijoittuminen toistensa suhteen.

R

Ramachandran plot *Ramachandran-kuvaaja, kiertokulmakartta, phi-psi-kuvaaja* Proteiinin aminohapot on esitetty kuvaajana, jossa x-akselilla on pääketjun phi-kulma ja y-akselilla vastaava psi-kulma. Kuvaajan avulla voidaan tarkastella proteiinin stereokemiaa ja sekundäärirakenteiden esiintymistä.

rare-cutter Restriktioentsyymi, jonka tunnistuskohtia on harvassa.

restriction endonuclease *restriktioentsyymi, restriktioendonukleasi* Bakterofageja bakteereissa pilkkovia entsyymejä. Restriktioentsyymit tunnistavat tietyn, yleensä palindromisen, DNA-sekvenssikohdan ja katkaisevat kaksijuosteisen DNA:n tästä kohdasta. Käytetään yleisesti molekyylibiologisissa tekniikoissa.

reverse transcriptase *käänteiskopioija tai -transkriptaasi* Retrovirusten DNA-polymeraasientsyymi, joka pystyy tuottamaan DNA-sekvenssin RNA-sekvenssin perusteella. Käytetään yleisesti molekyylibiologisissa tekniikoissa.

S

salt bridge *suolasilta* Proteiinin rakenteen sisäinen ionisidoksen kaltainen vuorovaikutus positiivisesti ja negatiivisesti varautuneen aminohapon välillä.

supersecondary structure *supersekundäärirakenne*
Säännöllinen, useammasta sekundaäri-
rakenneyksiköstä koostuva rakenne.

T

termination factor *lopetustekijä* Proteiini,
joka vaikuttaa transkription päätekohtan
tunnistamiseen RNA-synteesissä.

topoisomerase *topoisomeraasi* Entsyymi, jo-
ka muuttaa DNA:n kolmiulotteista muo-
toa vaikuttamalla superkierteisyyteen (su-
percoiling) katkaisemalla DNA:n toisen
tai molemmat juosteet.

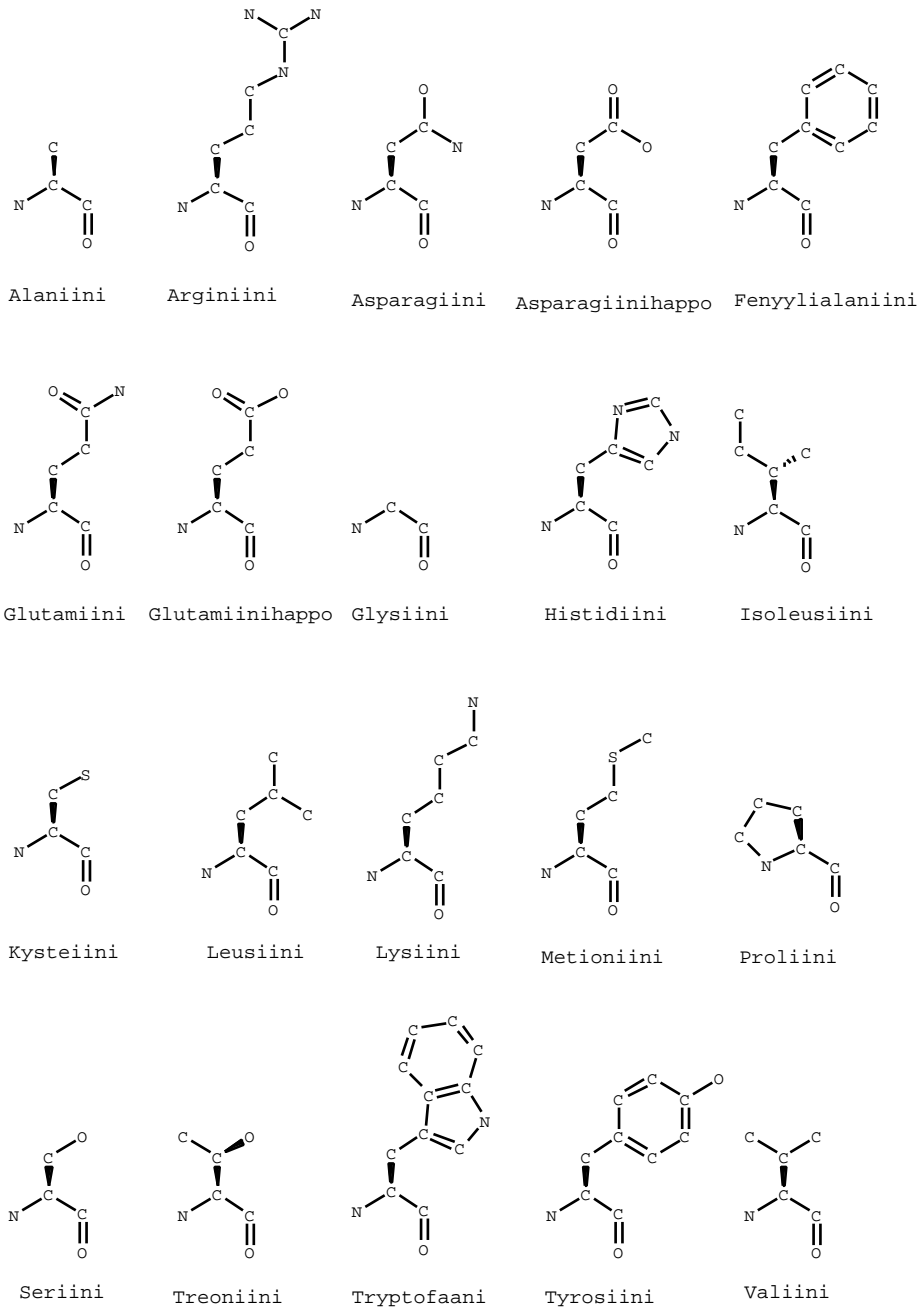
Z

zinc finger (metal-binding finger)

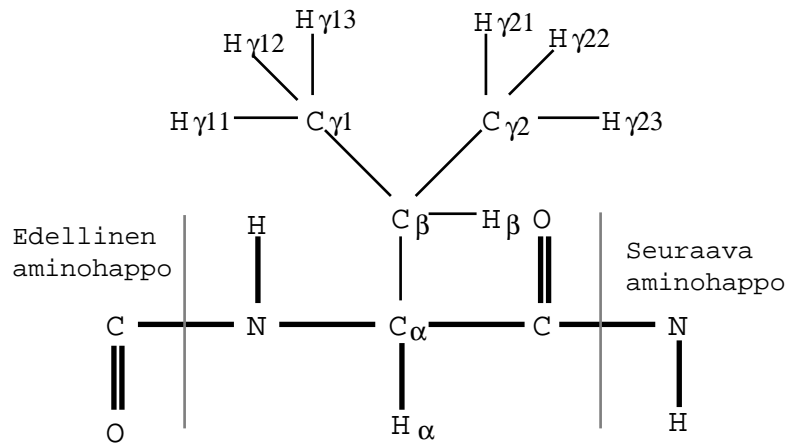
sinkkisormi Joukko sekvenssimotiiveja,
joille on yhteistä neljän histidiinin ja/tai
kysteiinisivuketjun avulla tapahtuva sinkki-
ionin sitominen. Sinkkisormet yleensä
tunnistavat ja sitovat nukleiinihappo-
pocketuja.

alanine	Ala	A	alaniini
arginine	Arg	R	arginiini
asparagine	Asn	N	asparagiini
asparagine or aspartic acid	Asx	B	asparagiini tai asparagiinihappo
aspartic acid	Asp	D	asparagiinihappo
cysteine	Cys	C	kysteiniini
glutamic acid	Glu	E	glutamiinihappo
glutamine	Gln	Q	glutamiini
glutamine or glutamic acid	Glx	Z	glutamiini tai glutamiinihappo
glycine	Gly	G	glysiini
histidine	His	H	histidiini
hydroxylysine	Hyl		hydroksilyysiini
hydroxyproline	Hyp		hydroksiproliniini
isoleucine	Ile	I	isoleusiini
leucine	Leu	L	leusiini
lysine	Lys	K	lyysiini
methionine	Met	M	metioniini
phenylalanine	Phe	F	fenyylialaniini
proline	Pro	P	proliini
serine	Ser	S	seriini
threonine	Thr	T	treoniini
tryptophan	Trp	W	tryptofaani
tyrosine	Tyr	Y	tyrosiini
valine	Val	V	valiini

Taulukko 2: Yleisimmät aminohapot ja niiden lyhenteet



Kuva 3: 20 yleisen aminohapon rakenteet ilman vetyatomeja.



Kuva 4: Aminohappojen atomien nimeäminen. Aminohappojen atomien nimeämisessä käytetään yleensä seuraavaa kaavaa: Pääketjun muodostavat aminoryhmä (N–H), α -hiili (C_α) ja siihen liittynyt vety (H_α), sekä karbonyyliryhmä ($C = O$). Sivuketjun atomit nimetään hiilestä pois päin siirryttäessä kreikkalaisin aakkosin ($\beta, \gamma, \delta, \epsilon, \zeta, \eta$). Vetyjen ja haarautuvien sivuketjujen kohdalla käytetään lisäksi numerointia erottamaan atomit toisistaan. Alla olevassa esimerkissä on nimetty valiinin kaikki atomit. Esimerkissä pääketjun atomit on lihavoitu.

5 Algoritmit, ohjelmat ja tietokannat

A

AMBER Biologisille makromolekyyleille kehitetty mallitusohjelmisto ja voimakenttä. <http://www.amber.ucsf.edu/amber/amber.html>

B

BLAST (Basic Local Alignment Search Tool) Ohjelma/algoritmi joka etsii hakusekvenssin kanssa samankaltaista sekvenssiä ja rinnastaa sekvenssit paikallisesti. Hakusekvenssi ja -tietokanta voivat olla nukleotidi- ja proteiinisekvenssien mikä tahansa yhdistelmä eli DNA-sekvenssillä voi hakea myös aminohapposekvenssitetokannasta ja päin vastoin. <http://www.ncbi.nlm.nih.gov/BLAST/>

BLITZ Rinnakkaislaskentaa hyödyntävä Smithin ja Watermanin algoritmiin pohjautuva sekvenssihakuehjelma.

BLOCKS Tietokanta ja analyysipalvelin hyvin samankaltaisten proteiinisekvenssi-alueiden etsintään ja vertailuun. <http://www.blocks.fhcrc.org/>

BLOSUM (Blocks Substitution Matrix) Pisteytysmatriisisarja, jota käytetään sekvenssianalyysissä. Perustuu BLOCKS-tietokannan kaukaista sukua olevien sekvenssien paikallisiin rinnastuksiin.

bootstrap *itselataava* Menetelmä, jossa algoritmin antamaa tulosta testataan toistamalla algoritmia sekoitetuilla aloitusarvoilla tai syöttötiedoilla. Sekoitus tehdään satunnaisesti, mutta kuitenkin tavalla, jonka ei pitäisi vaikuttaa käytetyn menetelmän antamaan tulokseen.

byte (MB) *tavu* Määrämittainen, yleensä kahdeksan bitin muodostama kokonaisuus. Esimerkiksi yhden kirjainmerkin esittämiseen tarvitaan yleensä yksi tavu.

C

CHARMM Biologisille makromolekyyleille kehitetty mallitusohjelmisto ja voimakenttä. <http://yuri.harvard.edu/>

Chou-Fasman analysis *Choun ja Fasmanin analyysi* Yksi proteiinien sekundaärirakenteen ennustusmenetelmistä.

Clustal (ClustaW, ClustalX) Monen sekvenssin (DNA- tai aminohappo-) rinnastusohjelma. http://www.infobiogen.fr/docs/ClustalW/clustalw_help.html

consensus sequence *konsensussekvenssi* DNA-, RNA- tai aminohapposekvenssiryhmän perusteella matemaattisin menetelmin tehty sekvenssi, joka kuvastaa nukleotidien tai aminohappojen yleisintä muotoa kussakin sekvenssin kohdassa. Alueet, joilla vastaavuus on suuri, heijastavat yleensä yhteistä toiminnallista historiaa.

covariance *kovarianssi* Kahden muuttujan yhteistä vaihtelua kuvaava tunnusluku.

D

Dali Menetelmä ja WWW-palvelin, jolla voi verrata proteiinin 3D-rakennetta PDB-tietokannassa oleviin rakenteisiin. Katso myös FSSP. <http://www.ebi.ac.uk/dali/>

DDBJ (DNA DataBank of Japan) Japanilainen nukleotidisekvenssitetokanta, jota ristiinpäivitetään EMBL:n ja GenBankin kanssa. <http://www.ddbj.nig.ac.jp/Welcome.html>

Discover Kaupallinen, MSI:n tuottama molekyyli-simulaatio-ohjelmisto, joka soveltuu myös biologisten makromolekyylien simuloimiseen.

distance geometry *etäisyysgeometria* Molekyylien rakennemäärityssä käytettävä laskennallinen menetelmä. Menetelmässä määritetään joukko atomien välisten etäisyyksien välisiä raja-arvoja, joiden perusteella luodaan molekyylin rakennemalleja. Käytetään etenkin NMR-spektroskopian yhteydessä.

distance matrix *etäisyysmatriisi* Taulukko, joka kuvaa taksonien parittaisia eroja.

docking *telakointi* Molekyylien sovittami-

nen toisiinsa esimerkiksi ligandien sitoutumista tutkittaessa.

E

EM algorithm (expectation maximisation algorithm) *odotuksen maksimointialgoritmi* Yleinen menetelmä maksimaalisen samankaltaisuuden laskemiseen.

EMBOSS (The European Molecular Biology Open Software Suite) Sekvenssianalytiikan ohjelmistopaketti. Eurooppalaisten bio-ohjelmistokehittäjien yhteistyöhanke, jossa yhdistetään jo olemassa olevia ja kehitettäviä ohjelmia ja työkaluja julkisesti saatavaksi ohjelma- ja kirjastokokonaisuudeksi. <http://www.uk.embnet.org/Software/EMBOSS>

Ensembl EBI:n ja Sanger Centerin ylläpitämä genomidataan keskittynyt WWW-palvelin. <http://www.ensembl.org/>

evolutionary distance *evoluutioetäisyys* Taksonien väliset erot esim. molekyyllisissä ovat tarkasteltavissa stokastisin menetelmin, joiden avulla kullekin taksoniparille saadaan evolutiivista etäisyyttä kuvaava tunnusluku.

ExpASy (Expert Protein Analysis System) Proteomiikkapalvelin, jota ylläpitää Swiss Institute of Bioinformatics (SIB). <http://www.expasy.ch/>

F

FastA Sekvenssihakuohjelma, joka perustuu Pearsonin ja Lipmanin algoritmiin. <http://fasta.bioch.virginia.edu/>

Fitch-Margoliash method *Fitch-Margoliash -menetelmä* Evoluutiikassa: puukaavioiden luomisessa käytettävä heuristinen menetelmä.

FSSP (Fold classification based on Structure-Structure alignment of Proteins) Tietokanta, jossa PDB-rakenteet on luokiteltu 3D-rakenteidensa perusteella Dali-menetelmällä. <http://www.ebi.ac.uk/dali/fssp/fssp.html>

FTP (File Transport Protocol) Yleinen tiedostojen siirtokäytäntö tietokoneiden välillä.

G

gap aukko Aukko sekvenssirinnastuksessa.

GAP Ohjelma sekvenssien parittaiseen vertailuun.

GCG (The Wisconsin Package of the Genetics Computer Group, Inc.) Sekvenssianalytiikkaohjelmisto, joka sisältää ohjelmia DNA-, RNA- ja proteiinisekvenssien karakterisointiin, hakuun ja vertailuun. http://www.accelrys.com/products/gcg_wisconsin_package/index.html

GenBank Yhdysvaltalainen NCBI:n ylläpitämä DNA-sekvenssitietokanta, jota ristiinpäivitetään EMBL:n ja DDBJ:n kanssa. <http://www.ncbi.nlm.nih.gov/Genbank/index.html>

Genehunter Kytkeäanalyysiohjelma, jota voi käyttää sekä parametriseen että ei-parametriseen kytkeäanalyysiin. Ohjelmalla voi analysoida vain pieniä sukupuita, mutta useita geenimerkkejä voidaan käyttää samanaikaisesti. <http://linkage.rockefeller.edu/soft/gh/>

genetic algorithm *geneettinen algoritmi* Evoluutiomekanismeja matkiva optimointimenetelmä. Algoritmi perustuu ongelman kuvaamiseen ratkaisujoukkona, johon kohdistetaan valintaa, tekijäinvaihdunnaa ja mutaatioita.

GeneWise Algoritmi/ohjelma, jota käytetään genomisten DNA-sekvenssien etsinnässä ja analyseissä. Perustuu kätettyihin Markovin malleihin. <http://www.sanger.ac.uk/Software/Wise2>

global alignment *kokonaisrinnastus* Sekvenssien rinnastaminen niin, että sekvenssien samankaltaisuuksia pyritään löytämään koko pituudelta. Katso local alignment.

GROMACS Biologisille makromolekyyille kehitetty mallitusohjelmisto ja voimakent-

tä. <http://rugmd4.chem.rug.nl/~gmx/>

H

HCA (hydrophobic cluster analysis) katso hydrophobic cluster analysis.

helical wheel *kierrekaavio* Peptidis-ekvenssin esitystapa, jossa kuvataan sivuketjujen sijoittuminen α -kierteen eri puolille. Käytetään amfipaattisten kierteiden etsimisessä.

heuristic *heuristinen, kokemusperäinen* Algoritmien yhteydessä heuristisella menetelmällä tarkoitetaan kokemusperäistä ja/tai oletuksiin perustuvaa menetelmää.

HMM (Hidden Markov Model)*kätkeyty* *Markovin malli* Malli, jossa havaintoja mallitetaan piilevällä kerroksella, jolle on määritelty Markov-ominaisuus (mallin seuraava tila riippuu vain edellisestä tilasta).

HMMer Ohjelmapaketti, joka käyttää HMM-menetelmää sekvenssianalytiikassa. <http://hmmer.wustl.edu/>

Hopp-Woods Yksi proteiinien hydropaattisuuden ennustusmenetelmistä.

HTML (HyperText Markup Language.) WWW-sivujen kuvauskieli.

HTTP (HyperText Transfer Protocol.) WWW-palvelimien käyttämä yhteydenpitomenetelmä.

hydropathy plot *proteiinin hydropaattisuuden ennuste* Kuvaaja, jossa on esitetty proteiinisekvenssin vesipakoisuus tai vesihakaisuus sekvenssin funktiona. Yleensä arvo lasketaan käyttämällä tietyn mitausta tarkastelualuetta ja (painotettua) keskiarvoa, jolloin myös sekvenssissä tutkittavan aminohapon ympärillä olevat kohdat tulevat huomioiduiksi.

hydrophobic cluster analysis (HCA)*hydrofobinen klusterianalyysi* Ennustaa hydropaattisuutta ja sekundäärirakenteita.

I

InsightII Kaupallinen molekyyli- mallitusohjelma, joka sisältää useita biologisten makromolekyylien mallitukseen soveltuvia ominaisuuksia. <http://www.accelrys.com/insight/index.html>

in silico tietokoneessa Toimenpide joka tehdään laskennallisesti. Esim. erilaiset simulaatiot.

J

Jukes-Cantor model *Jukesin ja Cantorin malli* DNA:n emäskorvautumismalli, jossa kaikki korvautumismahdollisuudet ovat yhtä todennäköisiä.

K

KEGG Kyoto Encyclopedia of Genes and Genomes Palvelu johon on koottu tietoa molekyylibiologisista reaktio- ja signaali- reiteistä sekä molekyylien ja geenien välisistä vuorovaikutuksista. <http://www.genome.ad.jp/kegg/>

Kimura model *Kimuran malli* 2-parametrinen ja 3-parametrinen Kimuran malli ovat yleisesti käytettyjä DNA:n emäskorvautumismalleja. 2-parametrisessa mallissa transitioiden todennäköisyys on eri kuin transversioiden todennäköisyys. 3-parametrisessa mallissa mallia on kehitetty siten, että eri transversioille käytetään kahta eri parametria.

ktup Monien hakualgoritmien (mm. FastA) parametri, joka määrittelee, kuinka pitkiä sekvenssialueita haun ensimmäisessä vaiheessa tarkastellaan. Parametri vaikuttaa sekvenssihaun herkkyyteen ja nopeuteen (arvo on tyypillisesti 1 tai 2 proteiineille, 3-6 nukleinihapoille; mitä pienempi, sitä herkempi, mutta hitaampi).

Kyte-Doolittle Algoritmi aminohaposekvenssin hydropaattisuuden laskemiseen.

L

Linkage ohjelma kytkentäanalyysiin. Ohjelma on käyttökelpoinen suurille sukupuille, mutta laskentalgoritmi mahdollistaa vain muutaman merkkigeenin samanaikaisen käytön. <http://linkage.rockefeller.edu/soft/linkage/>

local alignment paikallinen rinnastus Vertailtavista sekvensseistä etsitään huomattavan samankaltaisia alueita. Tulos esitetään yleensä taulukkona, jossa sekvenssien vastinkohdat ovat samassa sarakkeessa. Vert. kokonaisrinnastus, global alignment.

Logo Usean sekvenssin rinnastusohjelma.

M

MACAW Ohjelma, jolla voi paikallisesti rinnastaa sekvenssejä.

masking peittäminen, suodatus. poistaa yksinkertaiset toistot ja homopolymeeriset jaksot sekvenssistä ennen tietokantahakua tai linjausta, jotta ne eivät antaisi vääriä positiivisia tuloksia ja häiritسی hakua. Suodatusohjelmien (esim. SEG proteiineille ja DUST nukleinihapoille) peittämät aminohapot esitetään usein rinnastuksissa X-kirjaimina ja emäkset N-kirjaimina

maximum likelihood suurimman uskottavuuden menetelmä Menetelmää käytetään mm. evoluutiikassa etsittäessä fylogeneettistä puuta, joka parhaiten toteuttaa asetetun evoluutiomallin.

maximum parsimony parsimoniamenetelmä, niukkuusmenetelmä Sukulaissuhteiden selvitysmenetelmä, jonka tulostama fylogeneettinen puu vähimmin mahdollisin muutoksin selittää aineiston.

Megabyte (MB) megatavu Tiedoston tai tietokonelevyn koon tai kapasiteetin mitta. Yksi megatavu sisältää 2 potenssiin 20 tavua eli noin miljoona tavua.

Metropolis algorithm Metropolis-algoritmi Satunnaisuutta käyttävä optimointite-

htävien ratkaisumenetelmä.

midnight zone pimeä alue Sekvenssejä vertailtaessa alue, jolla vertailun luotettavuutta ei voida arvioida.

Minimum Evolution (ME) minimievoluutio Menetelmä, joka etsii lyhimmän etäisyys ehdot täyttävän puukaavion.

molecular clock molekyylikello Hypoteesi, jonka mukaan molekyylien evoluutio on ajan suhteen vakio eli molekyylien eroista voidaan suoraan laskea niiden erkane misajankohta.

molecular dynamics molekyyliidynamiikka Simulointimenetelmä, jossa molekyylien liikettä seurataan ajan funktiona. Yleensä laskennassa käytetään klassisen mekaniikan liikeyhtälöitä.

molecular mechanics molekyylimekaniikka Menetelmä molekyylien rakenteen ja dynamiikan laskemiseen. Yleisnimitys laskennallisille menetelmille, joissa tutkitaan molekyylin avaruusrakenteen muutoksien vaikutusta potentiaalienergiaan. Yleisimmin käytetty molekyylimekaaninen tehtävä on voimakenttään perustuva komiulotteisen rakenteen optimointi.

molecular modeling molekyyli mallitus Molekyylin avaruusrakenteen tutkimus tai ennustus rakennemallien avulla. Tarkoittaa yleensä tietokoneavusteista molekyylien kolmiulotteisten mallien käsittelyä.

Monte Carlo method Monte Carlo -menetelmä Satunnaisuuteen perustuva numeerinen menetelmä.

most parsimonious tree (MPT) parsimonisin puu. Fylogeneettinen puu, joka selittää aineiston yksiköiden väliset erot vähimmin mahdollisin muutoksin.

MPT (most parsimonious tree) katso most parsimonious tree.

MultAlin Usean sekvenssin rinnastusmenetelmä. <http://www.toulouse.inra.fr/multalin.html>

MultiDisp Usean sekvenssin rinnastuksen visualisointiohjelma. <http://protein.uta.fi>

8080/cgi-bin/MultiDisp.cgi

multiple sequence alignment (MSA) *monen sekvenssin rinnastus* Sekvenssiaineiston esittäminen niin, että samankaltaiset kohdat on aseteltu päällekkäin.

multipoint linkage analysis *monipistekyt-kenttäanalyysi* Kyt kenttäanalyysi, jossa uuden lokuksen sijaintia kyt kenttäryhmässä (kromosomissa) tutkitaan usean sellaisen lokuksen avulla, joiden suhteelliset etäisyydet toisiinsa tiedetään.

N

NDB (Nucleic Acid Database) Nukleiinihapporakenteita sisältävä tietokanta. <http://ndbserver.rutgers.edu/>

Needleman-Wunsch algorithm *Needlemanin ja Wunschin algoritmi* Dynaamiseen ohjelmointiin perustuva sekvenssivertailualgoritmi kahden sekvenssin rinnastukseen. Algoritmi etsii kahden sekvenssin välisen parhaan mahdollisen kokonaisrinnastuksen.

neighbor-joining algorithm *neighbor joining -algoritmi, NJ-algoritmi* Evoluutiikassa: puukaavioiden luomisessa käytettävä heuristinen menetelmä.

neural network (artificial neural network) *neuroverkko, hermoverkko* Algoritmi, jonka toiminta jäljittelee tietojenkäsittelyä biologisissa hermoverkoissa. Neuroverkkoalgoritmien tarkka määritelmä vaihtelee eri yhteyksissä. Yleisiä tunnusmerkkejä ovat mm. 1. Tieto syötetään neuroverkkoon oppimisprosessin avulla, 2. verkko sisältää useita toisiinsa kytkeytyneitä yksikköjä, joiden välisiin vuorovaikutussuhteisiin tieto varastoituu.

nonparametric analysis *ei-parametrinen analyysi* Tilastotieteessä: ei oleteta mitään aineistoa kuvaavien jakaumien muodosta.

nonsense mutation DNA:n emäsjärjestyksen muutos, jonka seurauksena proteiinisynteesi pysähtyy. Stop-kodonin tuottava mutaatio.

nonsynonymous substitution *ei-syn-*

onyyminen substituuutio Aminohapon korvaus täysin toisentyypisellä aminohapolla.

NP-hard (Nondeterministic polynomial complete/hard) *NP-täydellinen tai -vaikea* Laskentaongelman ominaisuus. NP-täydellisille ongelmille tunnetaan vain hyvin hitaita ratkaisualgoritmeja, joiden aikavaatimus kasvaa eksponentiaalisesti algoritmin syöttötietojen määrän mukana (ns. kombinatorinen räjähdys). Tämän takia NP-täydellisten ongelmien ratkaisuun käytetään usein (epätarkkoja) heuristisia menetelmiä.

O

OMIM (Online Mendelian Inheritance in Man) Geneettisten sairauksien tietokanta. <http://www.ncbi.nlm.nih.gov/Omim/>

P

PAM (Point Accepted Mutation). Sarja pisteytysmatriiseja, joilla määritetään aminohappoja vertailtaessa, miten aminohapot vastaavat toisiaan eli mikä on niiden samankaltaisuuden aste.

parametric analysis *parametrinen analyysi* Tilastollinen menetelmä, joka tekee oletuksia populaation jakaumasta tarkastellun ominaisuuden suhteen.

parsimony katso maximum parsimony.

PAUP (Phylogenetic Analysis Using Parsimony) Fylogeniikkaohjelmisto, jonka avulla voidaan analysoida sekvenssien ja taksonien sukulaisuussuhteita useilla eri menetelmillä. <http://paup.csit.fsu.edu/>

PDB (Protein DataBank.) Makromolekyylien kolmiulotteisten rakenteiden tietokanta. <http://www.rcsb.org>

PDBSum PDB-tietokannasta johdettu tietokanta, jonka tietueet sisältävät rakenteen lisäksi yhteenvedon rakenteesta sekä luokitteluja ja analyysituloksia. <http://www.biochem.ucl.ac.uk/bsm/pdbsum/index.html>

Pfam proteiiniperhetietokanta. Protei-

init on jaoteltu sekvenssin mukaisesti perheisiin. Kullekin perheelle tietokanta sisältää joukon hyvin rinnastettuja sekvenssejä, joiden perusteella luodulla HMM-mallilla muita sekvenssejä voidaan verrata ja luokitella. <http://www.sanger.ac.uk/Software/Pfam/>

PHD Sekvenssiin perustuva proteiinirakenteen ennustusohjelmisto, joka ennustaa annetun sekvenssin sekundäärirakenteen, pintarakenteen sekä solukalvon läpäisevien sekvenssien sijainnin.

Phylip Fylogeniikkaohjelmisto, jonka avulla voidaan analysoida sekvenssien ja taksonien sukulaisuussuhteita useilla eri menetelmillä. <http://evolution.genetics.washington.edu/phylip.html>

PIR (Protein Identification Resource) Sekvenssitietokannoista johdettu tietokanta, joka on tarkoitettu protei-
inisekvenssien ja niitä vastaavien nukleiinihapposekvenssien tunnistamiseen ja analysointiin. <http://www-nbrf.georgetown.edu/pir/>

ProDom Tietokanta, joka sisältää automaattisesti luokiteltuja protei-
inidomeeniryhmiä. <http://www.toulouse.inra.fr/prodom.html>

profile alignment *profiilirinnastus* Sekvenssirin-structure-activity relationship nastus, jossa pisteytysmatriisiin sijasta tai lisäksi käytetään paikkakohtaista pisteytysmatriisia eli profiilia. Yleensä profiili tehdään tunnetun monen sekvenssin rinnastuksen perusteella. Profiilirinnastuksessa voidaan pisteytysmatriisiin perustuvaa rinnastusta paremmin ottaa huomioon tietyille sekvenssiperheelle tyypilliset piirteet.

Prosite Tietokanta, joka koostuu biologisesti merkittävistä sekvenssijaksoista ja -perheistä sekä niihin liittyvistä sekvenssimotiiveista ja -profiileista. <http://www.expasy.ch/prosite/>

PSI-Blast (Positio Specific Iterative BLAST) BLAST-ohjelman versio, joka tekee BLAST-haun perusteella löytyneistä sekvensseistä sekvenssiprofiilin, jonka

avulla ohjelma suorittaa uuden, tarkeman haun. Tulosten perusteella tehdään uusi sekvenssiprofiili ja haku niin kauan, kunnes merkittävästi samankaltaisia sekvenssejä ei enää löydy.

PubMed NCBI:n ylläpitämä bio- ja lääketieteen kirjallisuustietokanta. <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi>

Q

quartet-puzzling Heuristinen menetelmä, jonka avulla suurimman uskottavuuden menetelmää voidaan soveltaa suurillekin aineistoille. Menetelmässä aineistosta määritetään suurinta uskottavuutta vastaavat puut kaikille mahdollisille neljän taksonin ryhmille, joista parhaat yhdistetään.

R

RNAMOT Tietokoneohjelma RNA:n primääri- ja sekundäärirakenteiden analyysiin sekä RNA-motiivien etsimiseen. <ftp://merck.bch.umontreal.ca/pub/Rnamot/>

S

SAR (structure-activity relationship) katso structure-activity relationship.

SCOP (Structural Classification of Proteins) Tietokanta, jossa PDB:stä löytyvät proteiinirakenteet on ryhmitetty hierarkiseksi kokonaisuudeksi. <http://scop.mrc-lmb.cam.ac.uk/scop/>

scoring matrix *pisteytysmatriisi* Taulukko, jolla paikasta riippumattomissa sekvenssirinnastusmenetelmissä määritetään eri nukleotidi- tai aminohappoyhdistelmien vaikutus rinnastuksen laatua tai todennäköisyyttä kuvaavaan pisteytykseen. Katso PAM ja BLOSUM.

secondary structure *sekundäärirakenne* Proteiinissa paikallisesti esiintyvät säännölliset kolmiulotteiset rakenteet. Yleisimmät sekundäärirakenteet ovat α -kierre ja β -säie.

SeqLab GCG-ohjelmistopakettin X-yhtey-

den kautta käytettävä käyttöliittymä.

SeqWeb GCG-ohjelmistopakettien verkkoalaimen kautta käytettävä käyttöliittymä. Esim. CSC:n SeqWeb rekisteröidyille käyttäjille: <https://seqweb.csc.fi/gcg-bin/seqweb.cgi>

Simwalk Ohjelma sekä parametriseen että ei-parametriseen kytkentäanalyysiin. Toisin kuin Genehunteria, ohjelmaa voi käyttää myös suurille sukupuille.

sliding window *liukuva ikkuna* Sekvenssi-analyysissä: tarkastelualue. Sekvenssiä käydään läpi ikkunan pituisissa palasissa niin, että ensimmäinen ikkuna alkaa ensimmäisestä sekvenssikohdasta ja analyysin edetessä ikkunaa liu'utetaan kohta kerrallaan eteenpäin sekvenssissä.

Smith-Waterman algorithm *Smithin ja Watermanin algoritmi* Dynaamiseen ohjelmointiin perustuva sekvenssien rinnastusmenetelmä. Algoritmi etsii kahden sekvenssin väliset parhaat mahdolliset paikalliset rinnastukset.

Solar Kytkentäanalyysiohjelma, joka on suunniteltu määrällisiin ominaisuuksiin vaikuttavien geenien paikallistamiseen.

SOM (Self Organizing Map)*Itseorganisoituva kartta* Yksi hermoverkkomenetelmien sovellus.

SRS (Sequence Retrieval System) WWW-ympäristössä toimiva sekvenssihaku työkalu. Esimerkiksi: <http://srs.csc.fi:8002>

Staden Molekyylibiologisen aineiston analyysityökalupaketti. Erityisen sovelias automaattisekvenssaattoriaineiston käsittelyyn.

stringency *ankaruus, vaativuus* Molekyylibiologiassa kuvaa usein sekvenssikoettimen sitoutumisolosuhteita tai sekvenssi-analyysissä vaadittavien pisteiden määrää esim. tarkasteluikkunassa.

structure-activity relationship (SAR) *rakenne-aktiivisuus-vuorovaikutus* SAR-menetelmissä käytetään tilastollisia menetelmiä rakenteellisten tekijöiden sekä ligandin tai proteiinin aktiivisuuden

vastavuuden arviointiin.

substitution matrix *pisteitysmatriisi*

Taulukko, jolla paikasta riippumattomissa sekvenssirinnastusmenetelmissä määritetään eri nukleiini- tai aminohappoyhdistelmien vaikutus rinnastuksen laatua tai todennäköisyyttä kuvaavaan pisteitykseen. Katso scoring matrix.

sum of pairs score *parisummapisteitys* Monen sekvenssin rinnastuksessa käytettävä pisteitysmenetelmä, jossa sarakkeen pistemäärä lasketaan kaikkien sarakkeessa olevien emäs/aminohappoparien parittaisten vertailujen pisteitysten summana. Esim. kolmen sekvenssin, A B C, tapauksessa parisummapisteitys on $(A+B)+(A+C)+(C+B)$.

SWISS-PROT Proteiinisekvenssitietokanta, jossa on pyritty hyvin jäseneltyyn proteiinisekvenssien ja niihin liittyvän tiedon kuvailuun. Sekvenssin lisäksi SWISS-PROT-tietueet sisältävät tietoa mm. proteiinin toiminnasta, alayksiköistä sekä vastaavista mutatoituista proteiineista. <http://www.expasy.ch/sprot/>

T

TCP/IP (Transport Control Protocol/Internet Protocol)Yksi Internetin käyttämistä tiedonsiirtokäytännöistä.

twilight zone *hämärän alue* Käytetään sekvenssivertailussa kuvaamaan sekvenssien samankaltaisuutta, kun niiden identtisyys on alle 25 prosenttia. Tämänasteinen samankaltaisuus voi olla satunnaista.

U

UPGMA (Unweighted Pair Group Method using arithmetic Averages). Yksinkertainen puukaavion muodostus algoritmi. Menetelmää on käytetty mm. geneettisiin etäisyyksiin perustuvassa fylogeniikka-analyysissä.

V

Viterbi-algoritmi. *Viterbin algoritmi* Las-

kee todennäköisimmän polun kätkeyn Markovin mallin läpi. Viterbin algoritmia käytetään kätkeyn Markovin mallien yhteydessä parhaimman sekvenssirinnastuksen etsintään. Algoritmi muistuttaa sekä rakenteeltaan, että käytöltään paljolti dynaamisen optimoinnin (Needlemanin ja Wunschinin algoritmi) käyttöä pisteytysmatriisimenetelmien yhteydessä.

W

W2H GCG- ja EMBOSS-ohjelmistojen verkkoselainkäyttöliittymä. <http://www.w2h.dkfz-heidelberg.de/>

weight *painotus, paino* Bioinformatiikan algoritmeissa kerroin, jonka mukaan tiettyä tekijää painotetaan suhteessa muihin tekijöihin. Esimerkiksi profiilirinnastuksessa tiettyyn sekvenssikohtaan sidottu painoarvo.

X

XML (Extensible Markup Language), Yleinen standardiformaatti, jolla voidaan kuvata rakenteista aineistoa ja dokumentteja tekstimuodossa. XML on suunniteltu erityisesti helpottamaan rakeenteellisen aineiston siirtoa ja esittämistä WWW-järjestelmässä.

Xplor Ohjelma, jota käytetään biologisten makromolekyylien rakenteen hienonnuks- ja määrittelylaskuihin röntgenkristallografiassa ja NMR-spektroskopiassa.

Z

Z-score *Z-arvo* Yleinen sekvenssianalyysiohjelmien parametri, jota käytetään arvioitaessa tietokannasta löytyneen sekvenssin ja hakusekvenssin välistä suhdetta. Z-arvo kuvaa sitä todennäköisyyttä, jolla yhtä samankaltainen sekvenssi voidaan löytää sattumalta.

Z-value katso Z-score.