Journal of Limnology and Freshwater Fisheries Research
limnofish
http://www.limnofish.org

Biyoteknolojinin Güncel Uygulamalarının Su Ürünleri Genetik Alanında Kullanılması: Yeni Nesil Dizileme Teknolojileri

Münevver ORAL*

Recep Tayyip Erdoğan Üniversitesi, Su Ürünleri Fakültesi, Rize, 53100, Türkiye

ÖZ

Geride bıraktığımız elli yıllık süreçte DNA dizi bilgisinin belirlenmesine yönelik muazzam çaba gösterilmiştir. Geliştirilen teknikler sayesinde kısa oligonukleotidlerden milyonlarca nükleotidlik tüm genom dizilemelerini tek reaksiyonda okuyabilen platformlara geçilmiştir. Bu ilerlemeler, Yeni Nesil Dizileme (YND) teknolojilerinin piyasaya sürülmesi ile gerçekleşmiştir. Kullanılan yöntemler, temelde bir genomun indirgenmiş temsilini oluşturan rastgele kütüphaneler (RADseq, ddRADseq, 2bRADseq, CROPS ve RRL) ile belli bir bölgeyi hedef alan kütüphaneler (RNAseq) olmak üzere ikiye ayrılırlar. Örneklerin hazırlanma süreci kısaca, DNA dizisi çıkarılması hedeflenen türün genomunun restriksiyon ya da sonikasyon yöntemi ile parçalara ayrılarak bir DNA kütüphanesinin oluşturulması ve ardından yüksek üretim hacmine sahip dizileme ekipmanları ile yeni sentezlenen DNA parçalarının yüksek kapasitede (paralel olarak) dizilenmesi, takiben de tüm bu dizilerin bir araya getirilmesi (*assembly making*) şeklinde özetlenebilir. Bu derlemede, literatürde en fazla kullanılan ve restriksiyon temelli yöntemlerden olan RADseq ve ddRADseq yöntemleri odaklı örneklerin hazırlanması ve biyoinformatik analizleri ele alınmıştır. Ülkemizde potansiyeli henüz keşfedilmemiş olan YND teknolojilerinin su ürünleri genetik literatüründeki kullanım alanları: (i) referans genom haritaları oluşturma (fiziksel), (ii) genetik bağlantı haritalamaları (QTL haritalama), (iii) popülasyon genetiği ve filogeni, (iv) TNP chip dizaynında, (v) verifikasyon ve validasyon çalışmalarında, (vi) ıslah amaçlı genotipleme ile (vii) sürdürülebilir su ürünleri yetiştiriciliği ve çevresel etkinin en aza indirilmesi noktasında bilgilendirici genetik izlenebilirlik alt başlıklarında derlenmiştir.

Anahtar kelimeler: Balık genetiği, yeni nesil dizileme teknolojileri, biyoteknoloji

MAKALE BİLGİSİ

DERLEME

Geliş : 28.02.2018
Düzeltme : 01.08.2018
Kabul : 15.08.2018
Yayım : 27.12.2018

DOI:10.17216/LimnoFish.399545

***SORUMLU YAZAR**

munevver.oral@erdogan.edu.tr
Tel : +90 464 223 1425

Using current applications of biotechnology in aquaculture genetics: Next Generation Sequencing Technologies

Abstract: There have been enormous attempts for determining DNA sequences within the last fifty years. Advances in technology have enabled the shift from the sequencing of short oligonucleotides to whole genome sequencing of millions of bases within a single reaction. Such advances have been attained with the launch of Next Generation Sequencing platforms. Techniques used involve two fundamental sections as generating a reduced representation of the genome of interest through random fragmentation based libraries (RADseq, ddRADseq, 2bRADseq, CROPS ve RRL) and target specific libraries (RNA seq). Briefly, library preparation involves fragmentation of the genomic DNA to be sequenced by using restriction digestion or sonication and then sequence massively in parallel via high-throughput sequencers and make assembly of short fragments. In the present review, RADseq and ddRADseq, the most commonly used techniques of NGS in the literature, have been focused on an explanation of library preparation and bioinformatics analyses. The potential that NGS technologies hold has not been fully understood in our country yet, the applications in aquaculture genetics are: (i) reference genome projects (physical), (ii) genetic linkage mapping (i.e. QTL mapping), (iii) population genetics and phylogeny, (iv) SNP chip design, (v) verification and validation studies, (vi) genotyping for selective breeding as well as (vii) genetic traceability studies for sustainable aquaculture and minimize environmental impacts.

Keywords: Fish genetics, next generation sequencing technologies, biotechnology

Alıntılama
Oral M. 2018. Biyoteknolojinin Güncel Uygulamalarının Su Ürünleri Genetik Alanında Kullanılması: Yeni Nesil Dizileme Teknolojileri LimnoFish. 4(3): 192-204. doi: 10.17216/LimnoFish.399545

Giriş

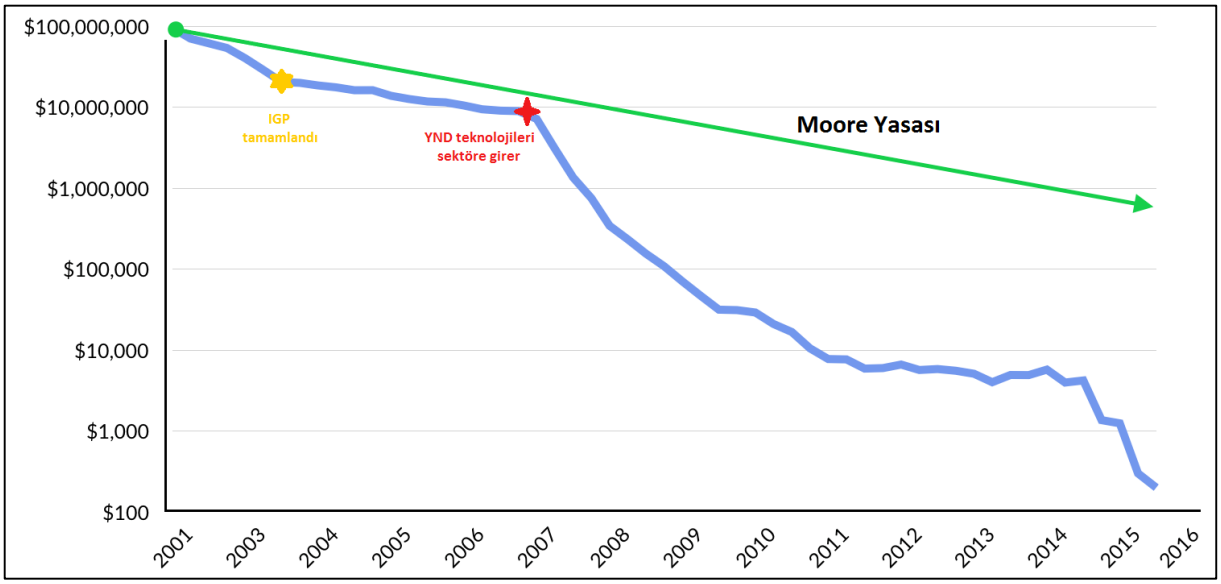
On dokuzuncu yüzyılın son çeyreğinde Gregor Mendel'in kalıtımın temel ilkelerini ortaya koymasını takiben 20.yy ortalarında DNA'nın keşfi ile başlayan süreçlerden bu yana bilim adamlarının başlıca merakı, gezegende bulunan tüm biyolojik yaşamın kodu olarak adlandırılan DNA molekülünün sırrını ortaya çıkarmak olmuştur. Şayet bu kodun şifreleri çözülebilirse yaşamın ilkelerine dair önemli bilgilerin elde edileceği şüphesizdir. 1990 yılında başlatılan İnsan Genom Projesi (İGP, Human Genome Project-HGP) tam da bu amaçla tasarlanmış ve hedeflenen proje teslim tarihinden 4 yıl önce 2003'te ilk taslağı erişime açılmıştır (IHGSC 2001; Venter vd. 2001). Bu sürpriz başarının mimarı, şüphesiz İGP ile ivme kazanan DNA dizileme sektörüne Yeni Nesil Dizileme (YND, Next Generation Sequencing-NGS) teknolojilerinin giriş yapması ve kısa zamanda yüksek çıktı hacmine sahip (high-throuhgput sequencing) veri setleri oluşturabilmesi ile sağlanmıştır.

DNA dizileme teknolojilerinin gelişimine bakıldığında 1977 yılı milat olarak alınır. Farklı yöntemleri temel alan ilk nesil dizileme teknolojilerinden olan (i) Maxam-Gilbert ve (ii) Sanger ve Coulson yöntemi aynı dönemlerde literatürdeki yerini almıştır. Maxam-Gilbert, 1977 yöntemi kimyasal temelli radyoaktif işaretlemeyi esas alır. Bu yöntemde dizilenecek DNA fragmanı denatüre edilerek tek zincirli hale getirilir ve fragmanlar 5' uçlarından ^{32}P radyoaktif bileşeni ile işaretlenerek restriksiyon enzimleri ile muamele edilir. Reaksiyon karışımında bulunan özellikli kimyasallar (dimetil sülfat ve hidrazin) DNA'yı istenilen bölgelerden keser. Örneğin pürin grubu bazlar için DNA fragmanında A ve G bulunan her noktada kesim yapılır. Aynı işlem bu kez pirimidin grubu bazları olan C ve T için de tekrarlandıktan sonra bu reaksiyonlar yüksek konsantrasyonlu poliakrilamit jelde yürütülmek suretiyle fragman büyüklükleri belirlenerek radyoaktif işaretleri görüntülenir. DNA dizisinin okunmasına jelin sonundan başlanır (elektroforez esnasında en küçük DNA fragmanı en hızlı ilerler prensibi dolayısıyla) ve üste doğru daha büyük DNA parçalarının dizilimleri yapılır. Ancak bu yöntem insan sağlığına zararlı yüksek oranda radyoaktif ve nörotoksik (hidrazin) kimyasallar içermesi ve de nispeten düşük çıktı hacmi ile (200-300 baz/reaksiyon) zaman içerisinde yerini Sanger ve Coulson dizilemeye bırakmıştır. Zincir sonlandırma ya da enzimatik yöntem olarak da adlandırılan (ii) Sanger ve Coulson yönteminde dizileme yapılacak DNA fragmanı önce restriksiyon endonukleazlar ile muamele edilerek küçük parçalara ayrılır ve her bir bazı temsil edecek şekilde dört tüp içerisinde denatürasyonu takiben oluşan tek zincirli kalıp DNA fragmanları, bu fragmanlara komplementer oligonükleoit diziler ile bu dizilerin uzamasını sağlayacak DNA polimeraz enzimi ve dideoksinukleotid karışımları kullanılır (Sanger ve Coulson 1975) – bu işlem günümüzde polimeraz zincir reaksiyonu (PCR) olarak bildiğimiz deney tüpü içersinde hedef DNA'nın çoğaltılmasını sağlayan reaksiyonun öncül formu ve fikir babasıdır. Bu yöntemin esası, kullanılan dideoksinukleotidlerin standart deoksinükleotidlere kıyasla DNA molekülünün uzamasını sağlayan fosfodiester bağı yapamamasından ileri gelir. Böylelikle merdiven basamağını andırır şekilde yeni sentezlenen DNA molekülleri kalıp DNA'dan ayrılarak elektroforeze tabi tutulur ve aynı Maxam-Gilbert (1977) yönteminde olduğu gibi büyüklüklerine göre ayrılır. Sanger ve Coulson yönteminde standart çıktı hacmi (200-500 baz/reaksiyon ile başlar ve sistem geliştirildikçe 1500 baz/reaksiyona kadar ulaşır) kimyasal yönteme kıyasla daha yüksektir ve zaman içerisinde yapılan düzenlemeler ile önceleri kullanılan radyoaktif işaretleyici zincir sonlandırıcı dideoksinukleotidler yerine daha güvenli floresan işaretçilerin kullanılmasıyla Sanger ve Coulson yöntemi hızla yaygınlaşmıştır. Geçmişten günümüze kullanılan DNA dizileme teknolojileri platformlarının tarihsel ve metodolojik gelişimi üzerine kapsamlı derlemeler yapılmış ve her platformun avantaj ve dezavantajları karşılaştırılmalı olarak rapor edilmiştir (Mertzker 2010; Mardis 2011, 2017; Guzvic 2013; Morey vd. 2013; Barba vd. 2014; Heather ve Chain 2016).

Geride bıraktığımız kırk sene içerisinde DNA dizileme teknolojileri çok farklı aşamalardan geçerek geliştirilmiştir ve daimi olarak kapasitesi arttırılarak ilerletilmiştir. Tüm bu süreçleri kontrol eden ana motivasyon ise çok daha kısa zaman içerisinde yaşamsal fonksiyonlar üzerinde karar alma noktasında önemli olan DNA dizi bilgisinin daha fazlasının çok daha ucuza, hızlıca sağlanabilir olması düşüncesi hakimdir. Bu emeklerin sonucu, geliştirilen hemen her platformda ürün (dizi bilgisi) artarken, fiyatlar her geçen gün hızla azalmıştır (Şekil 1). Öyle ki *Moore yasası** öngörüsünün bile ötesine geçmiştir (Muir vd. 2016). 2007-2008 yıllarında YND teknolojilerinin piyasaya sürülmesiyle (Illumina HiSeq X10, kısa fragman dizileme) az zamanda büyük ölçekli veri setleri üretilmiş ve bunun karşı konulmaz bir sonucu olarak gen bankasına gönderilen dizi bilgilerinde (özellikle tüm genom dizileme bilgilerinde) gözle görülür bir artışa sebep olmuştur (URL 1). Ortaya çıkan potansiyelin farkına erken varan biyoinformatik şirketleri, dizileme platformu tedarikçi firmaları ile iş birliği sağlayarak hızla gelişen bu alandaki eksiği gidermek ve sektörde marka olmak amacıyla yatırımlar yapmışlardır.

Bunlardan en dikkat çekeni Avrupa'da yer alan Edinburgh Genomics (URL 2)'tir. Sahip oldukları dizileme altyapısıyla NERC, BBSRC, MRC, Welcome Trust gibi kuruluşlarında direk ARGE desteği almakla birlikte; başta tıp (NHS – Ulusal Sağlık Sistemi; 100K genom projesi) olmak üzere ziraat, su ürünleri ve biyoloji alanlarında ulusal ve uluslararası organizasyonlar ile proje partnerlikleri yürütmektedir.

*:*Moore yasası – bir kanun olmaktan ziyade – 1965 yılında dünyanın en büyük teknoloji şirketlerinden olan Intel'in kurucusu Gordon Moore'un şirket ve piyasa gözlemleri doğrultusunda ortaya attığı deneysel bir bakış açısıdır. Mesajı ise gelişen piyasa ekonomisi ve teknolojisinde işletme gücünün her geçen yıl aynı maliyetler doğrultusunda en az ikiye katlanması gerektiği yaklaşımıdır.*



*Bu süreçte iki önemli kilometre taşı: (i) İnsan Genom projesinin ilk taslağının erişime açılması ve (ii) Yeni Nesil Dizileme teknolojilerinin piyasaya sürülmesi ile ivme kazanan ve hızla azalan maliyetin şematik gösterimi. Moore yasası beklenen azalma oranını temsil eder.

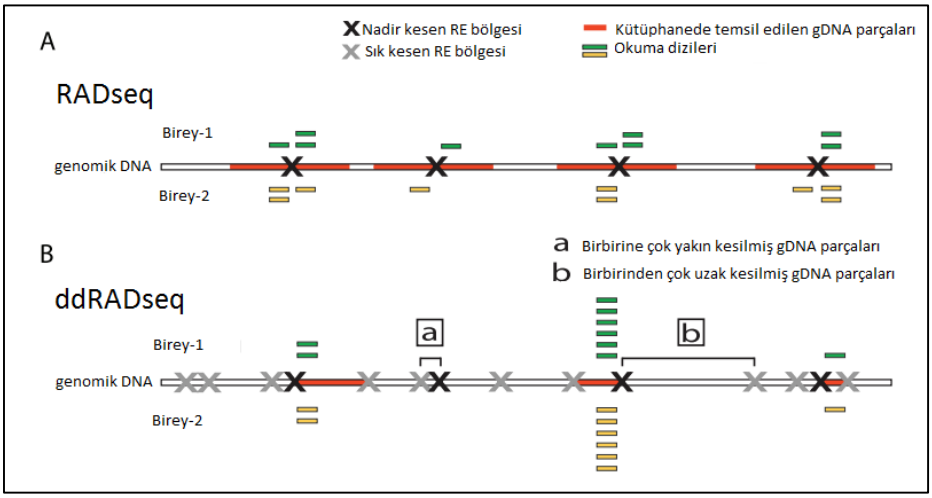
Şekil 1. DNA dizileme maliyetlerinin zamana bağlı değişim grafiği

Yeni Nesil Dizileme teknolojilerini diğer platformlardan ayıran en temel nokta aynı anda, çok büyük ölçekte ve paralel dizi bilgisinin eldesine imkân sağlamak üzere tasarlanmış olmalarıdır. Bu platformların yüksek-çıktı hacmi birkaç başlık altında toplanabilir: (1) Tek bir kütüphanede daha fazla birey genotiplenir (popülasyon ya da aile bazında), (2) birey başına daha fazla belirteç tanımlanır ve (3) her bir belirteç daha düşük hata oranı ve daha yüksek doğruluk oranları ile tespit edilirken paralel okuma derinliği (depth of coverage) de artar. Bu derinlik, genoma ait spesifik bir bölgenin kütüphane içerisinde kaç kez temsil edildiğinin bir göstergesidir. Proje amacına bağlı olarak çeşitlilik göstermekle birlikte standart bir YND kütüphanesinde en az 30 ve üstü birbirinden bağımsız fragman ile aynı bölgenin temsili, yüksek doğrulukta bir dizileme için gereklidir ve dizileme sonucu kalite kriterlerinden biri olarak kullanılır. YND teknolojilerinde tanımlanan belirteçler, hemen her genomda en yoğun bulunan özelikteki Tek Nükleotid Polimorfizmleridir (Single Nucleotide Polymorphism, SNP). Bu belirteçler, (i) yüksek çıktı hacimli veri setlerinin biyoinformatik analizinde kullanılan bir dizi yazılımının otomasyonuna maksimum uygunluk göstermeleri ve (ii) YND teknolojilerinin de yaygın kullanımına bağlı olarak büyük ölçekli genom analizleri (genome-wide analysis) ve haritalama (mapping) çalışmalarında geleceğin belirteci olarak gösterilmektedir (Liu 2011).

YND teknolojileri platformları çok çeşitli olmakla birlikte; yöntem farklılıkları, kütüphane oluşturma basamakları ve çıktı hacimleri yönünden temelde iki ana başlıkta toplanabilirler: (i) genomların indirgenmiş temsilini oluşturan ve DNA dizilemeyi içeren restriksiyon temelli kütüphaneler [**RADseq** (Baird vd. 2008; Etter vd. 2011), **ddRADseq** (Peterson vd. 2012), 2b-RADseq (Wang vd. 2012), CROPS (Van Orsouw vd. 2007), RRL (Van Tassel vd. 2008)] ve (ii) genomun belli bir bölgesini hedef alan (genelde transkripsiyon ve RNAseq temelli) yöntemlerdir. Bu derlemede literatürde en çok kullanılan yöntem olan RADseq (tek restiksiyon enzimi içerir) ve yine aynı yöntemden türetilmiş olan ddRADseq (çift restiksiyon enzimi içerir) tekniğine odaklanılmıştır.

RADseq yöntemini ddRADseq'den ayıran temel farklılık restiksiyon enzimi sayısı ve mekanik fragmentasyon (sonikasyon) kullanımıdır. Bu, aynı zamanda maliyeti de arttıran bir basamaktır. RADseq yönteminde tek restriksiyon enzimi ile muamele edilen örneklerin P1 ucundan adaptörlerin bağlanmasının ardından sonikasyona tabi tutularak istenilen boyuttaki DNA fragmanlarının seçilmesiyle P2 adaptörleri de bağlanarak dizilemeye hazır kütüphaneler oluşturulur. Dizileme sonucu yüksek çözünürlük ve derinlikte restriksiyon enzimi tanıma bölgesine komşu ve de çalışılan genom boyunca dağılım gösteren fragmanlar elde edilir (Davey vd. 2011). Restriksiyon enzimi kesim bölgesine komşu ve dizilemesi yapılan her bir fragman RAD gen bölgesi anlamına gelen, RAD lokus ya da (RAD-tag) olarak adlandırılır (Şekil 2). Bu tanımlanan RAD lokuslarının derinliği gerek TNP'lerinin belirlenmesi gerekse de genotiplemenin yüksek doğrulukla yapılmasına olanak sağlar (Robledo vd. 2017).

ddRADseq yönteminde ise (sözlüğe bakınız), metoda aynı zamanda adını da veren çift restiksiyon enzimi kullanılır (Şekil 2). Bu enzimlerden biri genom üzerinde *sık kesim* yapabilen (örneğin: *Sph*I; GCATG^C) daha kısa tanıma bölgesi içerirken ikinci enzim ise daha *nadir kesim* yapabilen (*Sbf*I; CCTGCA^GG) ve nispeten uzun tanıma bölgeleri içerir (Şekil 3). Gerek RADseq gerekse de ddRADseq protokolllerinde kullanılan restiksiksiyon enzimleri ve dolayısıyla elde edilmesi planlanan TNP belirteçleri kullanılan enzimin daha sık ya da nadir kesmesi seçilimi ile ayarlanabilir. Proje bütçesi, planlanan hedefler ve (şayet varsa) çalışılan türün genomuna ait veriler ışığında çalışma amaçları için en uygun enzim(ler)i bulabilmek için online programlar mevcuttur (Lepais ve Weir 2014). Burada kullanılan restriksiyon enzimlerini seçerken dikkat edilecek hususlar: (i) çalışılacak genomda kullanılacak dizileme teknolojisinin gereksinimlerine uygun fragmentasyon yapabilmesi ve (ii) birbirlerinin kesim noktalarını içermemesi, (iii) adaptörlere bağlantıyı sağlayacak 3' yapışkan uçlar vermesi ve (iv) her iki enzimin de aynı tampon çözelti içerisinde %100 aktivite gösterebilmeleri gereklidir (ddRADseq protokolünde tek bir reaksiyonda iki enzimli restiksiyon kesimi gerçekleştirileceği için bu basamak oldukça önemlidir). Örneğin, bilgisayar ortamında (*in silico*) gerçekleştirilen bir araştırmaya göre Gen Bankası'nda bulunan balık genomları üzerinde yapılan incelemeyle, yukarıda örneği verilen enzim kombinasyonu (*SbfI&SphI*) Illumina dizileme stratejisine optimum uyumlu nitelik ve nicelikte 300-600 nükleotidlik dilimde 2,500-4,000 arası fragman verdiği bildirilmiştir (Taggart ve Bekaert, kişisel iletişim). Yeni nesil dizileme teknolojilerinin su ürünleri ve balıkçılık alanında uygulamaları üzerine derlemeler literatürdeki yerlerini hızla almaktadır (Li ve Wang 2017; Kumar ve Kocour 2017; Robledo vd. 2017).



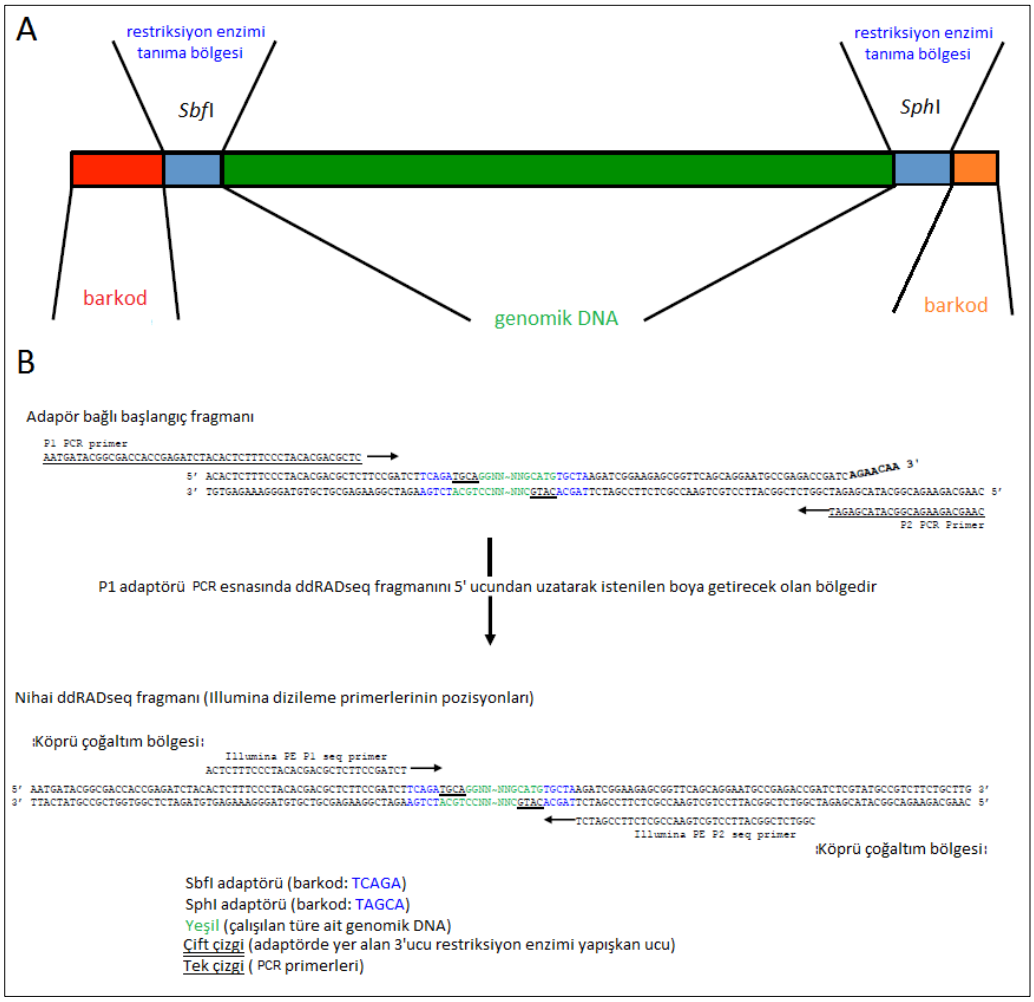
*(A) RADseq yöntemi tek bir restiksiyon enzimi ile kesimi rastgele fragmentasyon ile birleştirerek daha geniş ölçekte ve de bireyler arası daha dağınık örneklemeler yaparken (kırmızı ile işaretlenen DNA bölgeleri) (B) ddRADseq yönteminde kullanılan iki enzim her zaman aynı bölgelerden kesim yapacağından bireyler arası örneklenen gen bölgelerinin daha fazla birey arasında paylaşılmasına olanak sağlar ve böylelikle verimliliği arttırır. Şekil Peterson vd. 2012'den uyarlanmıştır.

Şekil 2. RADseq ve ddRADseq kütüphaneleri şematik gösterimi.

1. Restriksiyon temelli kütüphane oluşturma (ddRADseq odaklı)

YND laboratuvarında örneklerin hazırlanması süreci *kütüphane oluşturma* olarak adlandırılır ve kullanılan model organizmanın adı ve metodu özetleyecek şekilde isimlendirilirler. (Örneğin: Ssalar_RADtags,NileTipalia_sexQTL_ddRADtags). Kütüphane başına dizilenecek bireyleri arttırmak multipleksleme adı verilen ve dizilenecek DNA fragmanların önüne (P1) ve sonuna eklenen (P2)

adaptör kombinasyonlarının çeşitlendirilmesi ile mümkündür. Bu adaptörlerin her biri restiksiyon enzimleri ile uyumlu yapışkan uçlar içerir: Örneğin P1 adaptörü *SbfI* enzimi ile P2 adaptörü *Sph*I ile uyumlu uçlarının hemen ardından, barkod adı verilen 5 ya da 7 nükleotidlik birbirinden farklı dizileri içerir. Barkodlar her bir fragmana eklenen moleküler kimlik bilgileri olarak tanımlanabilirler. Kütüphane içerisindeki her bir bireye ait dizileme bilgisinin hatasız bir şekilde tanımlanabilmesine olanak sağlayan bu barkodları, kullanılacak dizileme platformuyla hibritleşmesini sağlayacak 20-40 nükleotidlik primer dizisi takip eder. Böylelikle dizileme cihazında okunacak ilk nükleotid barkod olacaktır ve onu sırasıyla restriksiyon enzimi yapışkan ucu (dizilenecek gDNA parçasıdır), çalışılan türe ait genomik DNA takip edecektir (Şekil 3).



Şekil 3. (A) ddRADseq kütüphanesinde dizilenen her bir fragmanın şematik görünümünü temsil ederken, (B) ddRADseq kütüphanesinde kullanılan adaptörlerin tasarımı ve adaptörler bağlandıktan sonra oluşan ddRADseq kütüphanesi fragmanlarının yapısını (5' barkodlar her iki adaptörde de mavi renkle kodlanmıştır) gösterir.

Böylelikle P1 okumaları için TGCA nükleotid dizisi, P2 okumaları için ise CATG nükleotid dizisi ile başlayacak olan fragmanlar birbirinden kolaylıkla ayırt edilebilir ve devam eden biyoinformatik veri işleme süreçlerinde barkodlar kolaylıkla dizilenecek gDNA fragmanlarından ayrılabilirler. Barkodları içeren adaptörlerin gDNA fragmanlarına bağlanmasını takiben tüm örnekler karıştırılarak bu aşamadan sonraki süreçler tek bir deney tüpü içerisinde sürdürülür. YND kütüphanelerinin en önemli basamaklarından bir tanesi istenilen fragman büyüklüğünün seçilimi aşamasıdır. Bu basamak bir dizi pürifikasyon (restiksiyon ve ligasyon reaksiyonları dolayısıyla ortamda fazla miktarda bulunan enzimleri uzaklaştırmak amaçlı) aşamasını takiben agaroz jelde ya da bu işlem için özel tasarlanmış cihazlarda (Pippin Prep PIP0001, Labgene scientific, İsviçre) gerçekleştirilir. Gerek aynı kütüphane içerisinde gerekse de farklı veri setlerini karşılaştırmak amaçlı farklı kütüphanelerde tanımlanan TNP belirteçlerinin yinelenebilirliği (repeatability) açısından aynı bölgenin kesimi büyük önem arz eder (Andrews vd. 2016). Kütüphane içerisinde bulunan adaptörleri (P1+P2) bağlanmış fragmanların niteliğini arttırmak için PCR kullanılır. Bu reaksiyon için adaptör bağlı generik primerler

kullanılır ve hazırlanan mastermix mümkün olduğunca fazla reaksiyona bölünerek PCR esnasında oluşabilecek sapmalar (*bias*; G ve C nükleotidlerince zengin gDNA bölgeleri ve kısa fragmanlar daha fazla çoğaltılma eğilimindedir) minimize edilir. Hazırlanan kütüphanenin PCR ile zenginleştirilmesinin ardından agaroz jelde kontrolü yapılarak seçilen fragman büyüklüğü ve miktarı tayin edilir. PCR esnasında kullanılan mastermix fazlasını da ortamdan uzaklaştırmak adına bir dizi pürifikasyonu takiben nihai kütüphane için seyreltmeler yapılır. Kütüphane son olarak ısı şoku (98°C, 2dk inkübasyon) ya da kimyasal denatürasyona (NaOH ile oda sıcaklığında 5dk inkübasyon) tabi tutularak YND cihazına yüklenir. Kullanılacak platforma göre farklılık göstermekle birlikte YND teknolojileri DNA dizileme işlemi yaklaşık 16-48 saat arasında sürer (örn: Illumina MiSeq platformu v2 kiti: 300 nükleotidlik çift yönlü okuma [paired-end read] sonucu 15 milyon fragmana eşdeğer yaklaşık 5Gb boyutunda veri seti üretirken, DNA dizileme işlemi 24 saat sürer).

2. Ham dizileme verilerinin kalite kontrol aşamaları

YND cihazlarının hemen hepsi dizileme sürecinde cihazda oluşmuş olan sorunlara yönelik rapor verirler; bu, kalite kontrolün ilk aşamasını oluşturur. Ancak söz konusu bilgiler daha teknik konuları kapsamakla birlikte genetik materyalde (yani kütüphanede) oluşan sorunlarla ilgili bilgilendirici değildir. Bu amaçla bağımsız bir yazılım olan FastQC (Babraham Bioinformatics) YND platformlarının hemen her çeşidinin veri çıktısını kullanarak hızlı ve basit bir dizi* kalite kontrol analizleri gerçekleştirir. Bu aşama veri analizlerine geçmeden önceki en temel ve önemli basamaktır. Zira yüksek hata oranı tespit edilen kütüphanelerden sağlıklı bir genotip bilgisi elde edilemeyeceği için proje amacından sapmalar gözlemlenebilir.

YND teknolojilerinin gelişimine paralel oluşan bu verilerin analizini yapacak birçok farklı yazılım ve platformu da beraberinde getirmiştir. Bu derlemenin amacı, var olan tüm platformlara ilişkin kapsamlı bir anlatımı tercih etmektense, verilerin biyoinformatik analizini içeren genel bir taslak sağlamak ve en çok kullanılan; aynı zamanda düzenli olarak güncellenen platformlara ilişkin değerlendirme yapmaktır. Konu ile ilgili detaylı okuma yapmak isteyenler için Mandoiu ve Zelikovsky (2016) ve Liu (2017) kaynaklari yön verici olacaktır.

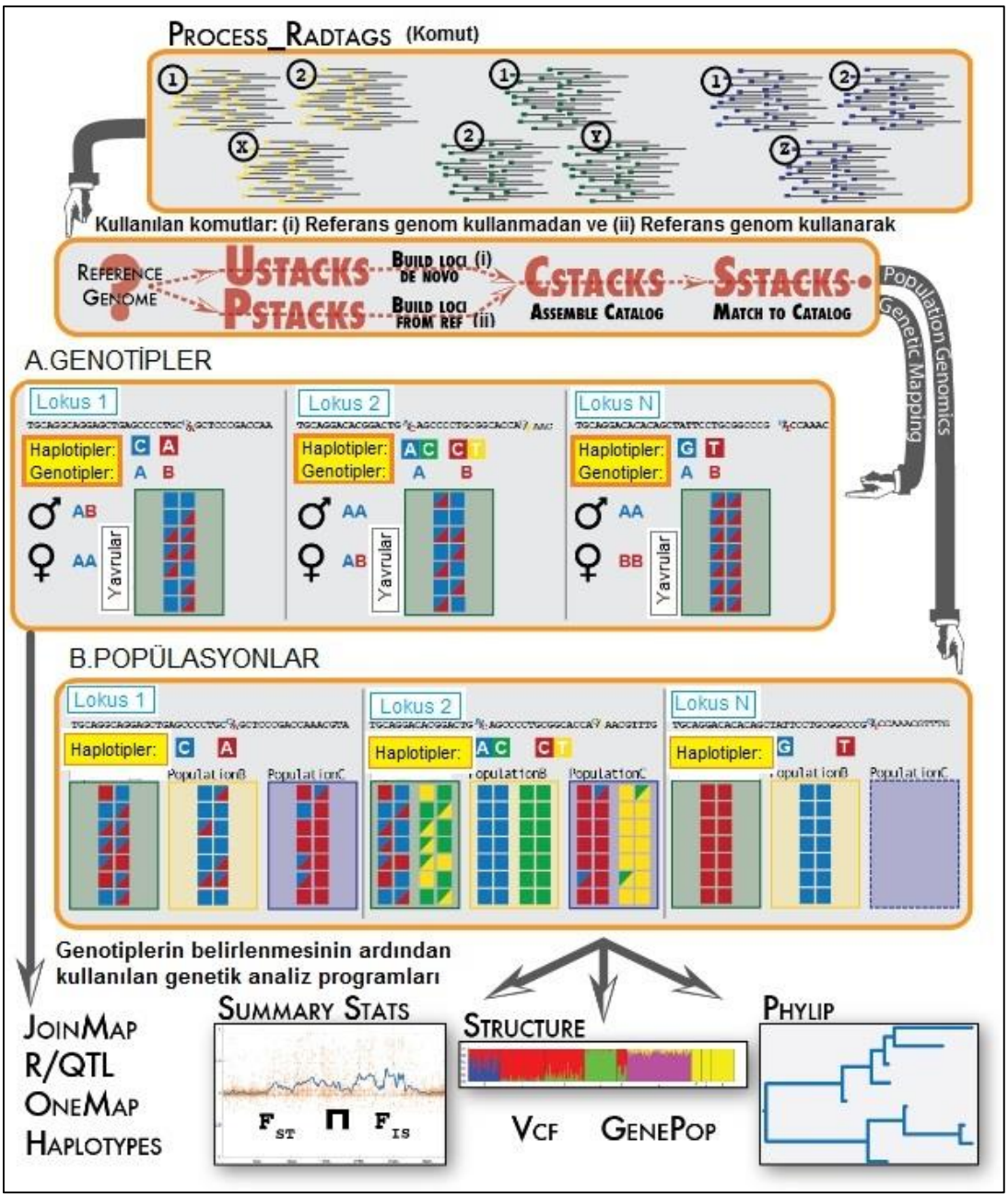
**:FastQC temel istatistiki verilerden kütüphanede yer alan spesifik kontaminasyonların içeriğine kadar toplam 12 analiz modülünden oluşur. Program yazılımı gereği dizileme verisinin normal dağılım gösteren formatta olmasını bekler ve her analiz için basit bir kaldı/geçti testi uygular. Ancak dizilenen YND kütüphanelerin amaçları ve kuruluş şekilleri göz önüne alındığında bazı analizlerde sapmalar gözlemlemek olasıdır. Örneğin, RADseq kütüphanelerinin baş kısımlarında adaptörlerin bulunduğu bölgelerde nükleotid kullanım sıklıklarında gözlenen sapmalar normaldir ve beklenir. Dolayısıyla bu raporun sonucunun kütüphane yapılış biçimi ve verilerin doğası göz önüne alınarak dikkatli bir şekilde değerlendirilmesi gerekir.*

3. Verilerin biyoinformatik işlenmesi

3.1. Küçük parçalardan büyük resme (dizilerin birleştirilmesi-*making assembly*)

YND teknolojilerinin büyük ölçekli verilerinin bir araya getirilerek anlamlı bütünler oluşturulması işlemi dizilerin birleştirilmesi anlamına gelen parçaları birleştirme (*making assembly*) olarak adlandırılır. Bu süreç özellikle Illumina teknolojisi gibi kısa okumalar yapan platformlarda büyük önem arz etmektedir. Dizilerin birleştirilmesi iki şekilde yapılır: (1) *de novo* assembly ve (2) *referans genom tabanlı* birleştirme işlemleridir. *De novo* birleştirme çalışılan türe ait hiçbir genetik bilginin olmadığı koşullarda ilk basamak olarak kullanılır. Latince kökenli bir kelime öbeği olup, *en başından* anlamındadır. Bu durum özellikle model tür statüsünde olmayan ve birçoğunun genetiğine dair henüz bilgi kazanılmamış sucul türler için YND teknolojilerinin en önemli avantajıdır. Referans genom tabanlı dizi birleştirme işleminde ise dizileme cihazından elde edilen veriler öncelikle çalışılan türün genomuna (gen bankasında bulunması durumunda) hizalanmasının ardından genotipler belirlenir. Dizilerin birleştirilmesi esnasında en sık kullanılan yazılımların başında Stacks (Catchen 2011, 2013) ve daha yakın zamanda geliştirilmiş olan PyRAD (Eaton 2014) ve dDocent (Puritz vd. 2014) yer alır. Maksimum benzerlik istatistiki yöntemi tabanlı çalışan Stacks; dizi varyantları ve polimorfizmi dizileme hatalarından ayırır. Gerek aile (QTL haritalama genetik kroslarında) gerekse de popülasyon analizlerinde güvenle kullanılır. İki ana modülde çalışan Stacks yazılımda ilk aşama: (i) Temizleme ve de-multipleksleme (*process_radtags*) olarak adlandırılırken ikinci aşama (ii) lokusların oluşturulması ve genotiplerin belirlenmesidir (*ustacks/pstacks, cstacks, sstacks*) (Şekil 4). İlk kez dikence balığında (üç dikenli balık, *Gasterosteus aculeatus;* Hohenlohe vd. 2010) RADseq tabanlı popülasyon genetiği çalışmasında kullanılan Stacks yazılımı o tarihten bu yana 48.versiyonuna yükseltilmiştir (URL 3). Sonrasında geliştirilen

dDocent yazılımının (Puritz vd. 2014) üç ayrı deniz balığı türünde (*Sciaenops ocellatus*, *Lutjanus campechanus* ve *Lutjanus vivanus*) aynı verileri kullanarak yaptığı karşılaştırmalı analizler sonucu dDocent'in Stacks'a kıyasla daha fazla TNP tanımladığını rapor etmesine rağmen Stacks (Catchen, 2013; URL 4) bugün hala restriksiyon tabanlı kütüphanelerden RADseq ve ddRADseq verilerinin biyoinformatik analizlerinde literatürde en fazla kullanılan yazılım olarak karşımıza çıkar.



Şekil 4: Stacks yazılımının iki ana aşamada yürütülen biyoinformatik analizlerinin şematik gösterimi: (1 aşama) process_radtags modülü ile temizleme ve de-multipleksleme işlemleri ve ardından (2.aşama) küçük parçaların bir araya getirilerek (making assembly) dizilerin birleştirilmesi ve genotiplerin belirlenmesi işlemleri. Stacks klavuzundan alınmıştır (URL 4).

4. YND teknolojilerinin su ürünleri genetik çalışmalarında kullanım alanları

Çok değil birkaç yıl öncesine kadar fazla sayıda genetik belirteç ile çalışmak pahalı olduğundan sadece ekonomik değeri olan ve en çok çalışılan birkaç tür için moleküler belirteç bilgileri erişilebilir konumdaydı. Ancak, gelişen dizileme teknolojileri ve hemen her geçen gün azalan maliyetler birçok sucul tür (gerek ekonomik gerekse de ekolojik ya da evrimsel önemi olan) için de yeni bir çalışma ortamı yaratmıştır. Hızlı bir şekilde binlerce hatta on binlerce belirtecin çalışılan türün genomunu kapsayıcı nitelikte veri setleri oluşturması daha önce denenmemiş birçok biyolojik hipotezin de araştırılmasına olanak sağlamıştır. Su ürünleri genetik çalışmalarında YND teknolojilerinin başlıca

kullanım alanları: (i) referans genom haritaları oluşturma (fiziksel), (ii) genetik bağlantı haritalamaları (kalitatif karakter lokuslarının belirlenmesine yönelik haritalama çalışmaları, QTL haritalama), (iii) popülasyon genetiği ve filogeni, (iv) TNP chip dizaynında, (v) verifikasyon & validasyon çalışmalarında, (vi) ıslah amaçlı genotipleme, (vii) sürdürülebilir su ürünleri yetiştiriciliği ve çevresel etkinin en aza indirilmesi noktasında bilgilendirici genetik izlenebilirlik çalışmaları olarak listelenebilir. (i) Referans balık genomları oluşturulan sucul türler: Morina balığı, *Gadus morhua* (Star vd. 2011); Pasifik istiridyesi, *Crassostrea gigas* (Zhang vd. 2012); zebra balığı, *Danio rerio* (Howe vd. 2013); gökkuşağı alabalığı, *Oncorhynchus mykiss* (Berthelot vd. 2014); Avrupa deniz levreği, *Dicentrarchus labrax* (Tine vd. 2014); Japon yılanbalığı, *Anguilla japonica* (Kai vd. 2014); sazan balığı, *Cyprinus carpio* (Xu vd. 2014b); turna balığı, *Esox lucius* (Rondeau vd. 2014); Nil tilapyası, *Oreochromis niloticus* (Brawand vd. 2014); Asya levreği, *Lates calcarifer* (Vij vd. 2016); Akdeniz midyesi, *Mytilus galloprovincialis* (Murgarella vd. 2016); kalkan balığı, *Scophthalmus maximus* (Figueras vd. 2016); Atlantik somonu, *Salmo salar* (Lien vd. 2016); kanal kedi balığı, *Ictalurus punctatus* (Chen vd. 2016) olarak sıralanabilir. (ii) Ekonomik karakterlerin tanımlanması amacıyla yürütülen QTL haritalama çalışmaları kapsamında Campbell vd. (2014) gökkuşağı alabalığında bakteriyel soğuk su hastalığı yapıcı ajan olarak bilinen *Flavobacterium psychrophilum* ve infeksiyöz hematopoetik nekrozis virus (IHNV) hastalık direnci ile bağlantılı genetik belirteçler tanımlamak ve bu belirteçleri kullanarak (marker assisted selection) bahsi geçen hastalıklara dirençli üretim popülasyonları oluşturmuşlardır. (ii) Palaiokostas vd. (2013) cinsiyet belirleme üzerine yaptıkları çalışmalarda fonksiyonel erkeklerin belirlenmesi ve böylelikle sektörün ihtiyacına yönelik tamamı dişi üretim popülasyonları oluşturmasına katkı sağlamak amacıyla Atlantik halibutunda cinsiyet belirleme bölgesiyle yakından ilişkili belirteçler tanımlamıştır. Aynı araştırmacılar, Avrupa deniz levreğinde çok daha karmaşık ve sıcaklık ile etkileşim içerisinde olan bir cinsiyet belirleme mekanizmasıyla karşılaşmışlardır (Palaiokostas vd. 2015). Larson vd. (2016) oluşturdukları yüksek sayıda belirteç içeren genetik bağlantı haritası ile Sockeye somonunda (*Oncorhynchus nerca*) sıcaklık toleransı ve büyüme üzerine birden fazla QTL tanımlamışlardır. (iii) Hohenlohe vd. (2010) RADseq tekniği ile dikence balığı (üç dikenli balık, *Gasterosteus aculeatus*) doğal popülasyonlarında türün iç sulara adaptasyonunda rol oynaması muhtemel bir dizi aday gen belirlemiş ve popülasyon yapısını tanımlamışlardır. Deniz anemonu (şakayık) *Nematostella vectensis*'in genetik çeşitlilik ve popülasyon yapısının tanımlamaya yönelik yürüttükleri bir çalışmada Reitzel vd. (2013) geleneksel belirteçlerin yapamadığını YND teknolojileri (RADseq) yardımıyla gerçekleştirmiştir. YND teknolojilerinin popülasyon genetiği ve filogeni alanında kullanımına ilişkin detaylar çeşitli derlemelerde derinlemesine tartışılmıştır (Davey ve Blaxter 2010; McCormack vd. 2013; Andrews ve Luikart 2014). (iv) Popülasyon içi polimorfizm seviyesinin belirlenmesi amacıyla belirteç sayısına göre adlandırılan (örneğin:10K=10,000 TNP) çipleri ise yoğun genotiplemenin rutin olarak kullanıldığı ıslah programlarında görmek mümkündür. Atlantik somonu (Houston vd. 2014; Yanez vd. 2016), kanal kedi balığı (Liu vd. 2014), sazan (Xu vd. 2014a) ve gökkuşağı alabalığı (Palti vd. 2015) gibi türlerde geliştirilmiştir. Çiplerin dizaynı ile ilgili çok daha fazla çalışma yürütülmesine rağmen validasyonları tamamlanmadan yayınlaştırılmadıklarından sucul organizmalar için tanımlanmış sınırlı sayıdaki TNP çipi hâlihazırda literatürdeki yerini almıştır. (v) Su ürünleri araştırmalarında önemli yer tutan izogenik balık hatlarının üretiminde uniparental (tek ebeveynli kalıtım) bireylerin validasyon & verifikasyonunda ve yeni nesillere genetik materyali inaktive edilen ebeveynden kalıtımın gerçekleşmediğinin kanıtlanması amacıyla YND teknolojileri kullanılmıştır (Oral 2016). Benzer bir çalışmada Hou vd. (2016) ürettikleri izogenik Japon pisi balığında (*Paralichthys olivaceus*) birinci ve ikinci nesil klonların verifikasyonunda YND teknolojilerinden faydalanmışlardır. (vi) Çipurada genom kompleksitesini restriksiyon temelli yöntemlerden 2b-RADseq ile azaltarak ıslah için genotipleme amacıyla gram negatif bir bakteri olan *Pasteurella* hastalık direncini genomik öngörüye dayalı hesaplama yoluyla araştırmış ve genomik öngörü modeli geleneksel BLUP (en iyi doğrusal sapmasız tahminleme) modelinden çok daha başarılı sonuç vermiştir (r = 0.38-0.46). (Palaiokostas vd. 2016). (vii) Balıkçılık yönetimi ve kaynakların sürdürülebilir kullanımı noktasında doğru politikalar oluşturmak üzere mersin balıklarında (*Acipenser gueldenstaedtii, A. persicus, A. baerii* ve *A. naccarii*) yapılan bir çalışmada RADseq ile tanımlanan TNP belirteçleri kullanılarak izlenebilirlik ve koruma çalışması yürütülmüş; gerek su ürünleri gerekse de balıkçılık yönetmeliklerinin uygulanmaya konulmasına yönelik öneriler sıralanmıştır (Ogden vd. 2013). Yukarıda listelenen çalışma alanlarından bağımsız olarak, YND teknolojileri genotip bilgisinin çıktıya dönüştürülebildiği hemen her çalışma ve amaca hizmet eder.

Referans balık genom projelerinin sayıları hızla artmaya devam etse de, yüksek belirteç sayısına sahip genetik bağlantı haritaları ve bu haritaların sahip olduğu rekombinasyon bilgileri olmaksızın büyük balık genomlarını tamamlamak neredeyse imkânsızdır (Davey ve Blaxter 2010). *De novo* referans genom çalışmaları parçalar (scaffold) seviyesinde başlatılarak bu parçalar bir araya getirildikçe kromozomal seviyelere ulaştırılır (Atlantik salmonu genomu üzerinde yürütülen bir çalışmada gen bankasında erişime açılan genomun scaffold seviyesindeki ilk [*Ssalar_v1*; ASM23337v1)] ve kromozomal seviyedeki dördüncü [Ssalar_v4; AGKD00000000.4] versiyonları arasında büyük çapta iyileşme olduğunu tespit etmişlerdir (Oral 2016, 5.kısım). *Ssalar_v1* Atlantik somon genomunun 1/6'lık versiyonunu içerirken, DNA dizileme bilgilerine genetik bağlantı, transkripsiyon ve RNA bilgilerinin eklenmesi ve dizileme derinliğinin de arttırılmasıyla türün genom büyüklüğü olan $3{,}4x10^9$ nükleotide en yakın versiyona ulaşılmıştır.) Tüm çabalara rağmen geliştirilen algoritmalar özellikle dizilerin birleştirilmesi noktasında oluşan boşluklarda ve ökaryotik genomlarda sıklıkla yer alan tekrar dizilerinin olduğu bölgelerde salt dizileme bilgisi ile doğru hizalama yapamaz. Bu süreçlerde sahip olduğu yüksek belirteç sayısı ile çalışılan genom boyunca dağılım gösteren ve belirteçlerin birbirlerine olan bağlantı ve rekombinasyon bilgilerine sahip genetik bağlantı haritaları; bu sebeple, referans balık genomlarının geliştirilmesinde kilit rol oynar. RAD temelli genetik haritalar ve tanımlanan binlerce belirteç kullanılarak yakın zamanda Avrupa deniz levreği, gökkuşağı alabalığı, Nil tilapyası ve kalkan balığında referans balık genomları güncellendi. Bir diğer önemli haritalama yöntemi olan ve sentromerlerin lokasyonu ile kromozomların rekombinasyonca aktif (telomerik) ve pasif (sentomerik) bölgelerinin belirlenmesini sağlayan gen-sentromer haritalamaları (G-C mapping) da referans genom projelerinin iyileştirilmesinde önemli rol oynamaktadır (Oral vd. 2017). Bu uygulama verifikasyon amaçlı belirteç dizaynı için kritik öneme sahiptir. Sayılar hızla artarken, yeni veri setleri var olan referans balık genomlarının da kalitesinin arttırılmasına ve fonksiyonel gen bölgelerinin belirlenmesine olanak sağlayacaktır.

Sonuç

DNA dizileme teknolojilerinde inanılmaz bir dönüşüm ve gelişim yaşadığımız son on yıllık süreç incelediğinde, gelecekte yapılabilecek çalışmalar üzerine tahminlerde bulunmak oldukça zordur. Ancak şüphesiz ki bu teknolojiler, çok daha fazla veriye ulaşma noktasında daha hızlı ve azalan bütçelerde hizmet vermeye devam ettiği müddetçe, önümüzdeki süreçlerde genetik bilimci tüm araştırmacıların bu teknolojileri içeren çalışmalar yapması kaçınılmazdır. Özellikle restriksiyon temelli kütüphanelerden RADseq ve türevi ddRADseq yöntemlerinin; farklı restriksiyon enzimlerinin tekli ya da kombine kullanımlarının sağladığı çok yönlülük, hemen her organizma için uygun çalışma şartları yaratmaktadır.

Gelişen teknolojilere modern genetik uygulamalarını adapte etme noktasında en büyük zorluk üretilen büyük veri setlerinin analizi ve biyoinformatik işlenmesi sürecidir. Kuşkusuz var olan yazılımların düzenli aralıklarla güncellenmesi önemli bir faktördür, ancak yeni geliştirilen programların da araştırmacılara kullanım kolaylığı sağlaması ve çalışılan aile(ler) ya da popülasyonlarda benzer filtreler uygulayarak farklı laboratuvarlarca da tekrarlanabilir sonuçlar üretmesi gerekir. En nihayetinde gelişen teknolojilere ayak uydurma noktasında multidisipliner çalışma ile çalışılan türün biyolojik özelliklerine hakim uzmanların, biyoinformatik analizlerde tecrübeli ekiplerle bir araya getirilmesi başarının da anahtarı niteliğindedir. Ülkemiz 2023 hedefleri dahilinde *teknolojideki gelişmelerin yakından takibi ve ARGE ile dışa bağımlılığın en aza indirilmesi ve ileri teknolojili katma değeri yüksek ürünler ya da bilimsel çalışmaları geliştiren, teknoloji transferine olanak sağlayan ve uluslararası standartların oluşumunda etkili küresel bir oyuncu olmanın* gerekliliği vurgulanmıştır (URL 5).

Bu derlemede bilimde yeni bir çığır açan YND teknolojilerinin gelişimi incelenmiş ve su ürünleri genetiğindeki başlıca kullanım alanları tanıtılmıştır. Bu bilgiler ekseninde hedefimiz yüksek kesinlik ve çıktı (dizi bilgisi) hacim potansiyellerini bünyesinde barındıran YND teknolojilerini ülkemize kazandırmak ve söz konusu alanda multidisipliner çalışabilen akredite laboratuvarlar oluşturmak olmalıdır. Kullanılan metotların optimizasyonunu takiben sektörün ihtiyaçlarına cevap oluşturacak popülasyonların üretimi sağlanabilir. Bu sayede, uzun vadede moleküler biyoloji ve genetik temelli bilimler başta olmak üzere, tıp, tarım, ziraat ve hayvancılık gibi birçok alanı kapsayıcı projeler üretilebilir ve rutin genotiplemenin kullanıldığı ulusal ölçekli ıslah projeleri yürütülebilir.

Sözlük

Adaptör: ***Adapter*** – Dizilenecek DNA fragmanlarına bağlanmış ve sentezleme yoluyla dizilemenin gerçekleştirileceği okuma plakalarına hibridizasyonu sağlayacak primer bölgesini içeren kısa oligonükleotid dizileridir.

Barkod: ***Barcode*** – Her bir birey için farklı olacak şekilde moleküler düzeyde birey ayrımı yapmak üzere tasarlanmış 5 ila 7 nükleotidlik dizilerdir.

Parçalar: ***Scaffold*** – *De novo* referans genom haritalamanın ilk basamağında küçük fragmaların arka arkaya sıralı bir şekilde eklenmesiyle oluşturulan parçalardır. Bir sonraki basamakta dizileme bilgisinin arttırılması ve parçalar arasındaki boşlukların mümkün olduğunca kapatılmasıyla oluşturulan scaffoldlar kromozomalara yükseltilir.

Yüksek Çıktı Hacimli Dizileme: ***High-Throuhgput Sequencing*** – Yeni Nesil Dizileme teknolojilerinin gelişimiyle paralel dizileme sonucu tek seferde binlerce, milyonlarca fragman veri setinin üretilmesine olanak sağlayan dizileme yöntemi ve çıktısıdır.

Örnek Kütüphanesi: ***Sample Library*** – Tercihen her iki ucundan adaptör bağlı DNA fragmanları koleksiyonu bir diğer deyişle, YND öncesi örneklerin hazırlanması sürecidir.

Belirteç: ***Marker*** – Genom üzerinde yeri bilinen DNA dizilerine denir (genom işaretçileridir). Belirteçler, bir türü tanımlamada kullanılabileceği gibi kromozom üzerindeki yeri ya da sorumlu geni henüz tespit edilmemiş gen bölge(lerini) tanımlamada da kullanılabilirler.

gDNA: ***Genomic DNA*** – Bir organizmanın hücrelerinde (çekirdek) sahip olduğu DNA içeriğinin tamamını kapsar.

Tek Nükleotid Polimorfizmi: *Single Nucleotide Polymorphism* – Genom üzerindeki bir bölgede tek bir nükleotid pozisyonunda meydana gelen polimorfizmi tanımlar.

RADseq: ***Restriction Site Associated DNA Sequencing*** – Restriksiyon bölgesi ile ilişkili DNA dizileme yöntemidir.

ddRADseq: ***Double-Digest Restriction Site DNA Sequencing*** – Çift enzimli restriksiyon bölgesi ilişkili DNA dizileme yöntemidir.

2b-RADseq: ***Type IIb Restriction Enzyme Based RADseq*** – Restriksiyon endonükleazlardan yalnızca Tip II enzimlerinin kullanıldığı restriksiyon bölgesi ile ilişkili DNA dizileme yöntemidir.

Polimorfik dizilerin kompleksitesinin azaltılması: ***Complexity reduction of polymorphic sequences (CROPS)*** – AFLP temelli bir YND teknolojisi kütüphane oluşturma yöntemidir.

RRL: ***Reduced Representation Library*** – Restriksiyon temelli YND teknolojisi kütüphane hazırlama yöntemlerinden biridir.

Çift yönlü okuma/dizileme: ***Paired-end read***:– Aynı DNA fragmanının iki yönlü dizilenmesi işlemidir.

Referans genom olmaksızın dizi birleştirme işlemi: ***De novo Assembly*** – Çalışılan türde genom seviyesinde hiçbir bilginin olmadığı durumlarda sadece DNA dizileme bilgisi kullanılarak dizilenen parçaların bir araya getirilmesi işlemidir. Herhangi bir organizmada referans genom oluşturma sürecinin öncelikli basamağıdır.

Referans genom temelli dizi birleştirme işlemi: ***Reference based assembly*** – DNA dizi bilgilerinin çalışılan organizmanın referans genomuna hizalanmasıyla dizilerin birleştirilmesi işlemidir.

QTL: ***Quantitative Trait Loci*** – Kantitatif karakter lokusları, ilgilenilen ekonomik karakterler (örn: hastalık direnç genleri, büyüme geni, cinsiyet belirleme geni vb.) ile yakın bölgelerde bulunan ve bu sebeple önemli ilgilenilen gen bölgesi ile yüksek istatistiki bağlantı (örn: birlikte rekombine olma) gösteren gen bölgeleridir. Çoğunlukla poligenik (çok genli) kalıtım özelliği gösterirler.

Gen-Sentromer haritalama: ***Gene-Centromere mapping*** – Kromozomlar üzerinde bulunan ve hücre bölünmesinde kritik rol oynayan sentromer pozisyonlarının belirlenmesini sağlayan haritalama yöntemidir. Çalışma prensibi genetik bağlantı haritalarına benzerlik gösterir ancak sadece özel kroslardan (mayoz yoluyla üretilmiş ginogenetik aileler) üretilmiş ailelerde uygulanabilir.

BLUP: ***Best Linear Unbiased Predictor*** – En iyi doğrusal sapmasız tahminleme anlamına gelen bu terim ıslah çalışmalarında popülasyonda istenilen özelliklerin gelecek nesillerde ortaya çıkarılmasını ya da hâlihazırdaki ekspresyonunu arttıracak damızlık adaylarının seçiliminde kullanılır.

Kaynaklar

Andrews KR, Luikart G. 2014. Recent novel approaches for population genomics data analysis. Mol Ecol. 23(7):1661-1667. doi: 10.1111/mec.12686

Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA. 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. Nat Rev Genet. 17(2):81-92. doi: 10.1038/nrg.2015.28

Baird N, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresco WA, Johnson E. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One. 3(10), e3376. doi:10.1371/journal.pone.0003376

Barba M, Czosnek H, Hadidi A. 2014. Historic perspective, development and applications of next-generation sequencing in plant virology. Viruses. 6(1):106–36. doi: 10.3390/v6010106

Berthelot C, Brunet F, Chalopin D, Juanchich A, Bernard M, Noël B, vd. 2014. The rainbow trout genome provides novel insights into evolution after whole-

genome duplication in vertebrates. Nat Commun, 5, 3657.
doi: 10.1038/ncomms4657

Brawand D, Wagner CE, Li YI, Malinsky M, Keller I, Fan S vd. 2014 The genomic substrate for adaptive radiation in African cichlid fish. Nature 513: 375–381.
doi: 10.1038/nature13726

Campbell NR, LaPatra SE, Overturf K, Towner R, Narum SR. 2014. Association mapping of disease resistance traits in rainbow trout using restriction site associated DNA sequencing. G3 (Bethesda). 4(12):2473–2481.
doi: 10.1534/g3.114.014621

Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH. 2011. Stacks: building and genotyping loci de novo from short-read sequences. G3 (Genes, Genomes, Genetics), 1(3):171–82.
doi: 10.1534/g3.111.000240

Catchen J, Bassham S, Wilson T, Currey M, O'Brien C, Yeates Q, Cresko WA. 2013. The population structure and recent colonization history of Oregon threespine stickleback determined using restriction-site associated DNA-sequencing. Mol Ecol. 22(11):2864-83.
doi: 10.1111/mec.12330

Chen X, Zhong L, Bian C, Xu P, Qiu Y, You X vd. 2016. High-quality genome assembly of channel catfish, *Ictalurus punctatus*. Gigascience. 5:39.
doi: 10.1186/s13742-016-0142-5

Davey JW, Blaxter ML. 2010. RADseq: next generation population gentics. Brief Funct Genomics. 9(5-6):416-23.
doi: 10.1093/bfgp/elq031

Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML. 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat Rev Genet. 12(7):499–510.
doi: 10.1038/nrg3012

Eaton DAR. 2014. PyRAD: assembly of denovo RADseq loci for phylogenetic analysis. Bioinformatics. 30(13):1844–1849.
doi: 10.1093/bioinformatics/btu121

Etter PD, Preston JL, Bassham S, Cresko WA, Johnson EA. 2011. Local de novo assembly of RAD paired-end contigs using short sequencing reads. PLoS One 6(4):e18561.
doi: 10.1371/journal.pone.0018561

Figueras A, Robledo D, Corvelo A, Hermida M, Pereiro P, Rubiolo JA vd. 2016. Whole genome sequencing of turbot (*Scophthalmus maximus*; Pleuronectiformes): a fish adapted to demersal life. DNA Res. 23(3):181-192.
doi: 10.1093/dnares/dsw007

Guzvic M. 2013. The history of DNA sequencing. J Med Biochem. 32(4):301–12.
doi:10.2478/jomb-2014-0004

Heather JM, Chain B. 2016. The sequence of sequencers: The history of sequencing DNA. Genomics. 107(1):(1-8).
doi: 10.1016/j.ygeno.2015.11.003

Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA. 2010. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. PLoS Genet. 6(2):e1000862.
doi: 10.1371/journal.pgen.1000862

Hou J, Wang G, Zhang X, Sun Z, Si F, Jiang X, Liu H. 2016. Production and verification of a 2nd generation clonal group of Japanese flounder, *Paralichthys olivaceus*. Scientific Reports. 6, 35776.
doi:10.1038/srep35776

Houston RD, Taggart JB, Cezard T, Bekaert M, Lowe NR, Downing A. 2014. Development and validation of a high density SNP genotyping array for Atlantic salmon (*Salmo salar*). BMC Genomics. 15: 90.
doi: 10.1186/1471-2164-15-90

Howe K, Clark MD, Torroja CF, Torrance J, Berthelot C, Muffato M, vd. 2013. The zebrafish reference genome sequence and its relationship to the human genome. Nature. 496:498–503.
doi: 10.1038/nature12111

IHGSC, International Human Genome Sequencing Consortium: Initial sequencing and analysis of the human genome. 2001. Nature. 409:860−921. PMID: 1123701.
doi: 10.1038/35057062

Kai W, Nomura K, Fujiwara A, Nakamura Y, Yasuike M, Ojima N. 2014. A ddRAD-based genetic map and its integration with the genome assembly of Japanese eel (*Anguilla japonica*) provides insights into genome evolution after the teleost-specific genome duplication. BMC Genomics. 15:233.
doi: 10.1186/1471-2164-15-233

Kumar G, Kocour M. 2017. Applications of next-generation sequencing in fisheries research: a review. Fish Res. 186:11–22.
doi: 10.1016/j.fishres.2016.07.021

Larson WA, McKinney GJ, Limborg MT, Everett MV, Seeb LW, Seeb JE. 2016. Identification of multiple QTL hotspots in sockeye salmon (*Oncorhynchus nerka*) using genotyping by-sequencing and a dense linkage map. J Hered. 107(2):122–133.
doi: 10.1093/jhered/esv099

Lepais O, Weir JT. 2014. SimRAD: an R package for simulation-based prediction of the number of loci expected in RADseq and similar genotyping by sequencing approaches. Mol Ecol Resour. 14(6):1314–1321.
doi: 10.1111/1755-0998.12273

Li YH, Wang HP. 2017. Advances of genotyping-by-sequencing in fisheries and aquaculture. Rev Fish Biol Fisheries. 27(3):535–559.
doi: 10.1007/s11160-017-9473-2

Lien S, Koop BF, Sandve SR, Miller JR, Kent MP, Nome T. vd. 2016. The Atlantic salmon genome provides insights into rediploidization. Nature. 533:200–205.
doi: 10.1038/nature17164

Liu S, Sun L, Li Y, Sun F, Jiang Y, Zhang Y. vd. 2014. Development of the catfish 250K SNP array for genome-wide association studies. BMC Research Notes. 7:135.
doi: 10.1186/1756-0500-7-135

Liu ZJ. 2011. Next Generation Sequencing and Whole Genome Selection in Aquaculture. Iowa, USA: Wiley-Blackwell 237 s.

Liu ZJ. 2017. Bioinformatics in Aquaculture. Oxford, UK: Wiley-Blackwell 591 s..

Mandoiu I, Zelikovsky A. 2016. Computational Methods For Next Generation Sequencing Data Analysis. New Jersey, USA: John Wiley & Sons. 464 s.

Mardis ER. 2011. A decade's perspective on DNA sequencing technology. Nature. 470:198–203. doi: 10.1038/nature09796

Mardis ER. 2017. DNA sequencing sechnologies: 2006-2016. Nat Protoc. 12(2):213-218. doi: 10.1038/nprot.2016.182

Maxam AM, Gilbert W. 1977. A new method for sequencing DNA, Proc Natl Acad Sci. U.S.A. 74, 560–564.

McCormack JE, Hird SM, Zellmer AJ, Carstens BC, Brumfield RT. 2013. Applications of next-generation sequencing to phylogeography and phylogenetics. Mol Phylogenet Evol 66(2):526-538. doi: 10.1016/j.ympev.2011.12.007

Mertzker ML. 2010. Sequencing technologies - The next generation. Nat Rev Genet 11:31–46. doi: 10.1038/nrg2626

Morey M, Fernandez AM, Castineiras D, Fraga JM, Couce ML, Cocho JA. 2013. A glimpse into past, present, and future DNA sequencing. Mol Genet Metab. 110 (1-2):3-24. doi: 10.1016/j.ymgme.2013.04.024

Muir P, Li S, Lou S. vd. 2016. The real cost of sequencing: scaling computation to keep pace with data generation. Genome Biology, 17(53). doi: 10.1186/s13059-016-0917-0

Murgarella M, Puiu D, Novoa B, Figueras A, Posada D, Canchaya C. 2016. A first insight into the genome of the filter-feeder mussel *Mytilus galloprovincialis*. PLoS One 11: e0151561. doi: 10.1371/journal.pone.0151561

Ogden R, Gharbi K, Mugue N, Martinsohn J, Senn H, Davey JW. vd. 2013. Sturgeon conservation genomics: SNP discovery and validation using RAD sequencing. Mol Ecol. 22:3112–3123. doi: 10.1111/mec.12234

Oral M. 2016. Insights into isogenic clonal fish line development using high-throughput sequencing technologies. [PhD thesis] University of Stirling, Scotland, UK, available online.

Oral M, Colléter J, Bekaert M, Taggart JB, Palaiokostas C, McAndrew BJ, Vandeputte M, Chatain B, Kuhl H, Reinhardt R, Peruzzi S, Penman DJ. 2017. Gene-centromere mapping in meiotic gynogenetic European seabass. BMC Genomics, 18:449. doi: 10.1186/s12864-017-3826-z

Palaiokostas C, Bekaert M, Davie A, Cowan ME, Oral M, Taggart JB. vd. 2013. Mapping the sex determination locus in Atlantic halibut (*Hippoglossus hippoglossus*) using RAD sequencing. BMC Genomics 14:566. doi: 10.1186/1471-2164-14-566

Palaiokostas C, Bekaert M, Taggart JB, Gharbi K, McAndrew BJ, Chatain B. vd. 2015. A new SNP-based vision of the genetics of sex determination in European seabass (*Dicentrarchus labrax*). Genet Sel Evol 47: 68. doi: 10.1186/s12711-015-0148-y

Palaiokostas C, Ferraresso S, Franch R, Houston RD, Bargelloni L. 2016. Genomic prediction of resistance to pasteurellosis in gilthead sea bream (*Sparus aurata*) using 2b-RAD sequencing. G3 (Bethesda) 6: 3693–3700. doi: 10.1534/g3.116.035220

Palti Y, Gao G, Liu S, Kent MP, Lien S, Miller MR. vd. 2015. The development and characterization of a 57K single nucleotide polymorphism array for rainbow trout. Mol Ecol Resour. 15: 662–672. doi: 10.1111/1755-0998.12337

Peterson BK, Weber J, Kay EH, Fisher HS, Hoekstra HE. 2012. Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. Plos One, 7:(5), e37135. doi:10.1371/journal.pone.0037135

Puritz JB, Hollenbeck CM, Gold JR. 2014. dDocent: a RADseq, variant-calling pipeline designed for population genomics of non-model organisms. Peer J 2: e431. doi: 10.7717/peerj.431

Reitzel AM, Herrera S, Layden MJ, Martindale MQ, Shank TM. 2013. Going where traditional markers have not gone before: utility of and promise for RAD sequencing in marine invertebrate phylogeography and population genomics. Mol Ecol. (11): 2953–2970. doi: 10.1111/mec.12228

Robledo D, Palaiokostas C, Bargelloni L, Martinez P, Houston R. 2017. Applications of genotyping by sequencing in aquaculture breeding and genetics. Rev in Aquacult. 1–13. doi: 10.1111/raq.12193

Rondeau EB, Minkley DR, Leong JS, Messmer AM, Jantzen JR, Von Schalburg KR. vd. 2014. The genome and linkage map of the Northern pike (*Esox lucius*): conserved synteny revealed between the salmonid sister group and the Neoteleostei. PLoS One 9: e102089. doi: 10.1371/journal.pone.0102089

Sanger F. Coulson AR. 1975. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. J Mol Biol. 94, 441–448.

Star B, Nederbragt AJ, Jentoft S, Grimholt U, Malmstrøm M, Gregers TF vd. 2011. The genome sequence of Atlantic cod reveals a unique immune system. Nature 477: 207–210. doi: 10.1038/nature10342

Tine M, Kuhl H, Gagnaire PA, Louro B, Desmarais E, Martins RS. vd. 2014. European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. Nat Commun. 5: 5770. doi: 10.1038/ncomms6770

Van Orsouw NJ, Hogers RCJ, Janssen A, Yalcin F, Snoeijers S, Verstege E, vd. 2007. Complexity Reduction of Polymorphic Sequences (CRoPS™): A Novel Approach for Large-Scale Polymorphism Discovery in Complex Genomes. PLoS ONE 2(11): e1172. doi:10.1371/journal. pone.0001172

Van Tassel CP, Smith TPL, Matukumalli LK, Taylor JF, Schnabel RD, Lawley CT, Haudenschild CD, Moore SS, Warren WC, Sonstegard TS. 2008. SNP discovery and allele frequency estimation by deep sequencing of reduced represantation libraries. Nat Methods, 5, 247-252.
doi: 10.1038/nmeth.1185

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG. vd. 2001. The sequence of the human genome. Science 291, 1304–1351.
doi: 10.1126/science.1058040

Vij S, Kuhl H, Kuznetsova IS, Komissarov A, Yurchenko AA, Van Heusden P. vd. 2016. Chromosomal-level assembly of the Asian seabass genome using long sequence reads and multi-layered scaffolding.
PLoS Genet. 12: e1005954.
doi: 10.1371/journal.pgen.1005954

Wang S, Meyer E, McKay JK, Matz MV. 2012. 2b-RAD: a simple and flexible method for genome-wide genotyping. Nat Methods. 9: 808–810.
doi:10.1038/nmeth.2023

Xu J, Zhao Z, Zhang X, Zheng X, Li J, Jiang Y. vd. 2014a. Development and evaluation of the first high-throughput SNP array for common carp (*Cyprinus carpio*). BMC Genomics.15: 307.
doi: 10.1186/1471-2164-15-307

Xu P, Zhang X, Wang X, Li J, Liu G, Kuang Y. vd. 2014b. Genome sequence and genetic diversity of the common carp, *Cyprinus carpio*. Nat Genet. 46: 1212–1219.
doi: 10.1038/ng.3098

Yanez JM, Naswa S, Lopez ME, Bassini L, Correa K, Gilbey J. vd. 2016. Genome-wide single nucleotide polymorphism discovery in Atlantic salmon (*Salmo salar*): validation in wild and farmed American and European populations. Mol Ecol Resour. 16: 1002–1011.
doi: 10.1111/1755-0998.12503

Zhang G, Fang X, Guo X, Li L, Luo R, Xu F. vd. 2012. The oyster genome reveals stress adaptation and complexity of shell formation. Nature 490: 49–54.
doi: 10.1038/nature11413

URL 1. 2018. https://www.ncbi.nlm.nih.gov/genbank/statistics/ (Erişim tarihi: 16.01.2018).

URL 2. 2018. https://genomics.ed.ac.uk/ (Erişim tarihi: 17.01.2018).

URL 3. 2018. http://catchenlab.life.illinois.edu/stacks/ tarihi (Erişim tarihi: 16.01.2018), son güncelleme tarihi: 29 Aralık 2017.

URL 4. 2018. http://catchenlab.life.illinois.edu/stacks/manual/ (Erişim tarihi: 16.01.2018).

URL5. 2018. https://www.tubitak.gov.tr/tubitak_content_files/vizyon2023/Vizyon2023_Strateji_Belgesi.pdf (Erişim tarihi:24.01.2018), sayfa 36.