What is the contribution of genetics to periodontal risk?
Loos, B.G.; Papantonopoulos, G.H.; Jepsen, S.; Laine, M.L.

Published in:
The Dental Clinics of North America

DOI:
10.1016/j.cden.2015.06.005

Citation for published version (APA):
The multi-causal etiology and complexity of periodontitis with emphasis on the genetic risk factors

B.G. Loos¹, G. Papantonopoulos², S. Jepsen³, M.L. Laine¹

¹ Department of Periodontology, Academic Centre for Dentistry Amsterdam (ACTA), University of Amsterdam and VU University, Amsterdam, the Netherlands
² Department of Mathematics, Center for Research and Applications of Nonlinear Systems, University of Patras, Greece
³ Department of Periodontology, Operative and Preventive Dentistry, Center of Dento-Maxillo-Facial Medicine, Faculty of Medicine, University of Bonn, Bonn, Germany

Address correspondence to
Bruno G. Loos, DDS, MSc, PhD
Professor and chairman
Dept of Periodontology
Gustav Mahlerlaan 3004
1081LA Amsterdam
The Netherlands
Phone +31-20-5980558
Email b.g.loos@acta.nl

Keywords: periodontology; periodontitis; aggressive periodontitis; chronic periodontitis; etiology; complexity; complex disease; multi-causality; genetics; candidate gene; GWAS; dental; cardiovascular disease; ANRIL; GLT6D1; COX-2; IL1; IL10;

Dental Clinics of North America
Abstract

This review addresses the multi-causal etiology of periodontitis, in which genetic factors play a role. The various proposed causes for periodontitis always play a role simultaneously, but the relative contribution of each of these, varies from case to case. In young individuals often with aggressive periodontitis (AgP) a stronger contribution from genetic factors is apparent, while in older individuals often with chronic periodontitis (CP), the relative contribution of the established subgingival biofilms (environmental factors) and life style factors (e.g. smoking, stress, diet), play a more dominant role in the phenotype of the disease. Nevertheless always some genetic susceptibility is present, for CP estimated at 25%. Periodontitis is therefore a complex disease, i.e. it behaves in a nonlinear fashion. Actually the disease progression rate fluctuates, where the disease sometimes moves into an aberrant state of host response and then swings back into a resolving state; in between, a settlement zone is present where essentially no differences are found for immunological parameters between cases with periodontitis and healthy controls. The genotype determines part of this fluctuation and the extent of it. The disease is polygenic, i.e. multiple genetic variants in multiple genes determine the phenotype, but there are individual and ethnic differences in the genes involved. We are still at the early stage to have identified the involved genes, in comparisons to other chronic diseases, we need to count on it that at least 100 causative genes across various global populations exist in AgP and CP. To date, the genetic variations firmly and repeatedly associated with periodontitis in some populations are found within the following genes: ANRIL, COX2, IL1, IL10, DEFB1, while many proposed periodontitis candidate genes have not been firmly proven or replicated.
Keypoints

Periodontitis is a multi-causal disease, with each of the causal factors playing a role but the relative contribution of these vary form case to case.

The disease behaves in a nonlinear fashion, with periods of aberrant host response to periods within an active disease resolving state.

To date only a few of the multitude of possible genetic factors for periodontitis are identified.

Outline

1. Introduction
2. Periodontitis is a complex disease
3. Periodontitis development and progression
4. Some theoretical background on genetics
5. How to identify periodontitis associated genetic markers
6. The best replicated genetic factors for periodontitis
7. Discussion and conclusions
8. References
1. Introduction

Periodontitis is a chronic inflammatory disease of the supporting tissues around the teeth, which results in irreversible periodontal attachment loss, alveolar bone destruction, subsequent tooth mobility and ultimately, if left untreated, tooth exfoliation. Traditionally we differentiate between two types of periodontitis: aggressive periodontitis (AgP) and chronic periodontitis (CP) (1). Severe periodontitis occurs in about 8-15% of the population (2, 3) depending on the definitions used for severe periodontitis and depending on the specific study population subjected to epidemiological studies. In countries with a high availability of dental care, with dental and health awareness, and with preventive measures available, the prevalence of severe periodontitis may be <10% (4), while in countries with no dental care, the prevalence can even be >15% (5). Recent studies suggest that almost half of the population suffers from mild to moderate periodontitis (6, 7). Nevertheless, severe periodontitis is a disease occurring only in a minority of populations (8-15%) (6, 8) and specific susceptibility factors play a role.

This chapter discusses the multi-causal etiology and complexity of periodontitis with emphasis on the genetic risk factors. It is based in part on a recent review (9).

2. Periodontitis is a complex disease

The current concept of the etiology of periodontitis is that it is a complex disease. Complexity of periodontitis means that it is a process involving multiple causal components (10, 11), which interplay with each other simultaneously. Complex systems are almost always nonlinear. Nonlinearity in complex systems means that the causes and effects are disproportional so that a small cause may result in a large effect or a large cause may result in a small effect, and that the disease progression rate fluctuates, or rather, can move from one state to another and back. Nonlinearity in periodontitis is seen in periodontitis by revealing heterogeneity in its clinical course; the latter is a common clinical finding by the many periodontal specialists who treat periodontitis patients.

There are several main causal risk factors, (i) the subgingival bacterial biofilm on both the tooth root surface and on the pocket epithelial lining (12), (ii) genetic risk factors and epigenetic modifications (9, 13, 14), (iii) life style related risk factors, such as smoking, stress
and poor diet or micronutrients intake (15-17), (iv) systemic disease, notably diabetes (18) and (v) others factors as of yet unknown, but possibly occlusal disturbances or fremitus, iatrogenic aspects (19) (Figure 1).

Figure 1. Periodontitis is a complex disease; multiple causal risk factors act simultaneously in the onset and progression of the disease. Several main causal factors play a role, i.e. environmental, (epi)genetics, life style, systemic diseases and others.

The five main causal components for periodontitis can be brought together into a pie chart, (adapted from (10, 20)). Figure 2 present a generic multi-causality model for periodontitis, where each of the five causal components have an equal contribution.
Figure 2. A generic multi-causality model for periodontitis, where 5 causes are playing a role simultaneously, genetic factors (blue), life style factors (red), environmental factors, i.e. bacterial biofilms (yellow), systemic disease (orange), other (unknown) factors (dark blue). For each individual periodontitis patient the relative contribution of the causal factors varies. Adapted from (10, 20)

However, it is important to note that the relative contribution of each of the causal factors varies from patient to patient (see Figure 3). In general, the older patients with CP are considered to have a major contribution from environmental factors and life style factors: many years of biofilm accumulation and unfavorable life style behaviors like smoking, poor diet, and no or irregular visits to dental professionals, have likely contributed most (Figure 3). On the other hand, periodontitis in younger patients, for example suffering from AgP, can be caused to a greater extent by genetic factors. It is generally accepted that genetic factors play an important role in the disease susceptibility of AgP (Figure 3) (13, 21, 22).

Figure 3. Genetics contribute more in relative younger patients with aggressive periodontitis (AgP) than in adults with chronic periodontitis (CP). Conversely, in relative older patients the microbiological risk factors and life style factors contribute the most to onset and/or progression, while genetics play a smaller role (9, 13).
Evidence for the role of genetics in periodontitis has been gained from population, family and twin studies (23-26). Studies in twins and especially in monozygotic twins are a strong and preferred model to study heritability of a disease. Michalowicz et al (24) studied 110 twin pairs (63 monozygotic and 33 dizygotic twin pairs) with CP and showed that among monozygotic twins a higher concordance of alveolar bone loss patterns was seen than in the dizygotic twins. The heritability estimates for alveolar bone loss among reared-apart monozygotic twins was 38% (24). For the early-onset form of periodontitis, often equated with AgP, there is only one study that has reported on the heritability in twins (27). The study consisted of a low number of twins (seven monozygotic and 19 dizygotic twin pairs at age 12-17 years) and therefore no clear conclusions could be made on the concordance rate of early-onset periodontitis. Because of a relatively low prevalence of AgP in the general population, it is very difficult to identify enough affected twins to provide sufficient statistical power to test the concordance of this disease phenotype. Nevertheless, because of strong familial aggregation, rapid progression and early onset of disease, it is clear that genetic factors play a large role in the disease susceptibility of AgP (Figure 3) (21, 22). Thus, the earlier periodontitis manifests itself, the greater the role of genetic factors, which is similar to other complex inflammatory diseases.

3. Periodontitis development and progression

The bacteria in the subgingival biofilms on tooth root surfaces and on pocket epithelial linings, as well as their toxic and antigenic products (e.g. endotoxin, bacterial metabolic components), initiate the periodontal inflammatory reactions and these lead to the enhanced recruitment of polymorphonuclear neutrophilic granulocytes (PMN), and other inflammatory cells into the gingival tissues (28). Subsequently, the recruited immune cells, in particular PMN, but also the activated pocket epithelial cells and the fibroblasts in the underlying matrix, release pro-inflammatory mediators including cytokines, prostanoids and proteolytic enzymes (29).

The type and severity of the periodontal inflammatory reaction to the dental biofilm is determined by genetic risk factors and the other causal factors named above. The disease-type of inflammatory reactions in the periodontal tissues, results in an infiltrated connective tissue, loss of periodontal ligament and resorption of alveolar bone. This destruction needs to be
regarded as collateral damage due to the inflammatory processes. Collagen and intercellular matrix degradation is thought to be caused by the activity of a large variety of proteolytic enzymes derived from fibroblasts and PMN. Moreover, enhanced osteoclastogenesis and osteoclast activation is initiated by cytokines and prostanoids, like interleukin (IL)-1 and prostaglandin (PG)-E2 respectively (29, 30), resulting in periodontal destruction.

The host response, i.e. the inflammatory reactions, varies from person to person, and is determined by genetics and other factors known to lower host responses, such as an unfavorable lifestyle and/or systemic disease such as diabetes (gene-life style interactions). The genetic blueprint of an individual is not a black and white deterministic causal factor. It is very dependent on the type and function of variants in multiple genetic loci: (i) multiple genes play a role (polygenic) (gene-gene interactions) and (ii) not all disease cases are having the same genetic risk variants and (iii) multiple, genetic variations are present at the same time. All these genetic variations essentially modify the host response, which is also not at a constant level. The host response can fluctuate or swing between states. The patient’s susceptibility to periodontitis (and many other complex chronic diseases), is determined by the complex interplay between the environment (i.e. bacteria) and the host. Here genetic variations may cause modifications in the immune system, and the immune system is also affected by life-style factors such as smoking, stress and diet, as well as certain systemic diseases, e.g diabetes (Figure 1). Furthermore, epigenetic changes of DNA and mutations during lifetime may modify individual’s susceptibility to periodontitis.

A depiction of the complex system of periodontitis can be attempted on the basis of the current understanding and interpretation of complexity (11). Undisturbed subgingival biofilms become maximum climax pathogenic communities as described in the literature (12), which are considered a state of stable equilibrium. In mathematical terms, the bacteria of the subgingival biofilms come into contact with the host defense elements through random walk and beyond a criticality in their interaction, the confrontation of bacteria to the host defense is driven out of equilibrium. Interestingly, it has been proposed that there are keystone pathogens that elevate community virulence and the resulting dysbiotic community targets specific aspects of host immunity promoting an aberrant immune response (31). At this point, self-organization (SO) of the host defense system occurs (32). SO represents a spontaneous arrangement of the system from a state of initially separated parts to a final state of joined and
tied up parts. As the term implies, SO is a complex system’s evolutionary process without any external factor imposed. SO happens in a space of scale invariance. Such a space is provided by periodontal ligament, which is a scale invariant self-similar object (33). SO is based on (i) amplification of disease activity by positive feedback, (ii) balancing of disease activity by negative feedback, (iii) amplification of random fluctuations (all three situations are obvious causes of nonlinearity) and (iv) on multiple local interactions of the system’s components. In other words, periodontitis can fluctuate or move between an aberrant immunological reaction state and a resolving state (resolution) (34, 35).

Fluctuations and chance are parts of a complex system’s behavior (mathematical term for host response) and random perturbations (“noise”) facilitate the system to make a series of attempts to explore all possibilities and finally make a choice and stabilize at a state for some period; this can be regarded as an order out of “noise”. A coherent global behavior appears as SO, e.g. the system behaves as AgP (aberrant immune reactions) or CP (minor disease progression or in resolving state, often no differences in immune reactions found between CP cases and healthy controls). The resultant emergent global behavior offers the ground for predictability out of the system’s chaotic (aperiodic, sensitive to initial conditions and therefore unpredictable) behavior (36). After SO occurrence we can no longer reduce the description of the system to one of its parts: all contributing parts (genetics, environment, life style, etc) are tied up into a collective host defense factor.

It is suggested in the literature that even the immune response level mounted at the early stage of gingivitis is the determinant factor of periodontitis progression and not the presence of specific bacteria known for their virulent properties (34, 35). The host response level can vary widely without affecting the SO, and there is no real specific value for the host response level for SO to occur.

Thus we can formulate the hypothesis that the host response against the subgingival microbiome is a function of host genetics, environmental and life style factors, systemic disease(s) and unknown factors. However, the outcome is not simply a summation. There are no clear-cut borders in the behavior of complex systems; there is always a mixture of regularity, chance and fluctuations. In the majority of the population, the host response is basically “normal” regarding the susceptibility to periodontitis. The host response swings like
a pendulum, from the aberrant immune response to the resolving state, while it always passes through a settlement or accommodation zone with little disease progression (Figure 4), and where the level of the host response is essentially normal, not different from immune parameters in healthy controls (37). In the light of recent epidemiological data which point that periodontitis is a pandemic disease (7, 8), we propose the paradigm that the host response is mostly normal and comparable to healthy controls. However as the system is always in some sort movement, it can swing into an aberrant immune function (one can observe this as an AgP case) or it can swing into a resolving or resolution phase (slow periodontal progression of many CP cases). Both the aberrant and resolution type host responses for periodontitis have been suggested in the literature (28, 34, 35).

Figure 4. A recently published mathematical model (37), suggests that there are two zones of host defense level that overlap. We indicate the specific zones as on one side an “aberrant” (red zone) resulting in a relative fast periodontal progression rate (exponential units, mean local lyapunov exponent) and on the other side a “resolving” (blue) zone with minor or minimal progression. The overlap represents an accommodation or settlement zone for the majority of periodontitis patients (supported by clinical data, where no differences have been found in mean values of immunologic variables among aggressive and chronic periodontitis (CP) cases and healthy controls (37)). Thus, we suggest that the host response of aggressive periodontitis (AgP) cases tends to “swing” into the aberrant zone, whereas the host defense for chronic periodontitis (CP) cases might fluctuate mainly in the settlement and resolving zone (on X-axis the level of host response of periodontitis patients compared to healthy controls, being x fold times more/less than the “normal healthy” level). Note, with this model, cases can fluctuate from CP to AgP and back also in the resolving zone. We can compare the host response as a pendulum of a clock.

4. Some theoretical background on genetics
Hundreds of thousands genetic loci can be identified in the human genome (i.e. the complete set of chromosomes). These are stretches of DNA on the chromosomes with an order, with a start and an end. Many of these genetic regions do not contain a gene proper (a genetic region coding for a protein), rather just regulatory elements or DNA with unknown function (originally called “junk DNA”). There are some 20,000 to 30,000 genetic loci which do contain one or several genes coding for proteins and having flanking regulatory regions.

Genes direct the production of proteins with the assistance of enzymes and messenger molecules. In humans, the genes are located on 23 pairs of chromosomes, 22 pairs of autosomal chromosomes and 1 pair of sex chromosomes (XX for females and XY for males). From each pair, one chromosome is inherited from the father and one from the mother.

In the chromosomes, DNA is arranged in a double helix: two polynucleotide chains are associated together by hydrogen bonding between the nitrogenous bases. The pairing of the two single stranded nucleotide chains is complementary: G pairs only with C, and A pairs only with T; these are called base pairs (bp). The sequences of these four nucleotides determines the information encoded in the DNA. Virtually every single cell in the body contains a complete copy of the approximately 3 billion DNA base pairs that make up the genome (38).

A genetic locus containing a gene is illustrated in Figure 5. This genetic locus consists of various parts. A promoter region, a sequence of nucleotides upstream of the coding region that contains specific sequences of nucleotides that are essential for the regulation and initiation or inhibition of the transcription of the coding region. Introns are often relative large stretches of non-protein coding nucleotides within the gene, which also often participate in the regulation of transcription of DNA into mRNA. Another function of introns may be help in the splicing events of mRNA before the translation step. Introns are surrounded by exons, which code for the sequence of amino acids of a protein. For the transcription and translation into a protein, the genetic code in such an exon, is read in groups of three nucleotides; each trinucleotidet sequence (triplet) is called a codon, which encodes a specific amino acid. The collection of known exons in our genome is called the exome.
Figure 5. Schematic drawing of a genetic locus containing a gene (length $63.5 \times 10^3$ base pairs) that codes for a protein (here in this example for the vitamin D receptor), being flanked by regulatory regions. The genetic locus contains exons (indicated by red arrows) and introns (indicated by blue arrows). The exons contain the proper DNA coding sequences for the amino acids making up the protein. The vertical “bar code” below the genetic locus indicates all known positions where 100s of single nucleotide polymorphisms (SNPs) may exist on basis of genome sequencing of 1000 individuals (see Figure 6). Any give individual may have on average a SNP every 1200 base pairs. At each genome position is then a given allele is present (major or minor allele). Many SNPs are inherited “per block”, they are linked (linkage disequilibrium).

Genes can be transcribed in alternative ways, where the regulatory regions play decisive roles; thus with this, each of the estimated 20,000 to 30,000 genes in the human genome code for an average of three proteins (38). Proteins make up all the body structures, and of course are at the base of forming the teeth, oral mucosa and the periodontium, carry signals between cells, code for the immune system, as well as they can be enzymes and can control biochemical reactions. If a cell’s DNA is mutated, an abnormal protein or abnormal protein quantities may be produced, which can disrupt the body’s usual processes and lead to a disease.

Nucleotide sequencing technologies determined the exact orders of the base pairs in the DNA. The international Human Genome Project finished the first working draft sequence of the entire human genome in the year 2000 (39, 40) and a first high-quality reference sequence of the human genome was successfully completed in 2003. In May 2006, the finished high-quality version of the sequences of all human chromosomes was published. It showed that the genomes of any two people are more than 99% identical, but variations between the individual genomes exist and differ on average in about one in every 1,200 bp. Differences in
individual bases are by far the most common type of genetic variation, and are known as *single nucleotide polymorphisms* (SNPs) (Figure 5 and 6).

![Interindividual DNA sequence variability](image)

*Figure 6. SNP = Single Nucleotide Polymorphism. SNPs occur at random and are mainly “neutral”, i.e. without any consequence. There are about 11 x 10⁶ SNPs/individual and 30 novel SNPs evolve per generation.*

5. How to identify periodontitis associated genetic markers

5.1 Candidate gene approach

In general, the genetic loci and the known genetic polymorphisms occurring in such loci, are chosen for *candidate gene association studies* in periodontitis based on their relation with immune responses and/or have previously been associated with other chronic inflammatory diseases such as rheumatoid arthritis, cardiovascular disease, Crohn’s disease, type 2 diabetes, systemic lupus erythematosus. The overlapping genetic risk factors for common chronic inflammatory diseases is called *pleiotropy* (41, 42).

Approximately 11 million SNPs are estimated to occur commonly in the human genome. All known common SNPs are listed in the catalog of common genetic variation, the HapMap, which was generated by the HapMap project and first published in the year 2005 (43). It describes the characteristics of the variants, where they occur in the DNA, and how they are distributed among people within populations and among populations (www.hapmap.org).

The catalog with the highest coverage of human genetic variation, obtained from population-based sequencing was published in 2010 by the “1000 Genomes Project” (www.1000genomes.org), and reported that each individual differs by >11 million often very rare SNPs, and that each generation >30 de novo mutations arise per individual (44).
The different variants at a specific genetic position are called *alleles*, and the individual’s personal unique collection of alleles in his/her genome makes up the individual’s *genotype*. Two or more alleles for a given position may exist in nature, and occur with different frequencies. The *minor allele frequency* (MAF) in a population is the proportion of the least frequent allele at a given chromosomal/genetic position and can range from 0-50%. Variants with a MAF >5% in a population are termed common variants. If the MAF of a variant ranges between 1-5% it is called a rare variant. Genetic variants with frequencies <1% are called mutations.

![Example of the prevalence of a given SNP](image)

*Figure 7.* Case control studies compare the frequency of selected SNPs, representing alleles, in two well-defined groups of non-related individuals: controls, who are either known to be unaffected, or who have been randomly selected from the population, and cases with periodontitis. These studies can be used to estimate the disease risk conferred by the allele, which is expressed by the odds ratio (OR). The OR is the ratio of allele carriers to non-carriers in cases compared with that in controls, which gives the increase in disease risk for carriers compared to non-carriers.

In candidate gene association studies, allele frequencies of selected SNPs are determined and compared between a well-defined group of cases and controls (Figure 7). Next to the statistical significance, the results are expressed with an odds ratio (OR). The ORs of common gene polymorphisms associated with complex diseases are typically ≤1.5 (in Figure 7 for the purpose of illustration an extreme OR of 3.7 is shown) (45).
A mutation or a genetic variant may have no effects or may have moderate to strong effects. For example, if a transition has taken place within the coding region of a gene, it may result in an amino acid substitution and therefore an altered protein structure, which may affect its function (nonsynonymous SNP). Or, when such mutations have taken place in the promoter region of the gene, it may alter the expression levels of the gene. Accordingly, genotypic differences among individuals can contribute to phenotypic variation, which is termed the genetic variance. How strongly a genetic variant affects the susceptibility to a disease is defined as the genotype relative risk (GRR), the ratio of the risk of disease between individuals with and without the genotype. A ratio of 1.1 equates to a 10% increase in risk and is often expressed as the odds ratio (OR). However, carriership of a genetic variant or mutation does not inevitably lead to disease, but only a proportion of individuals with a mutation or risk variant will develop the disease, this is known as penetrance. The severity of the disease in individuals who carry the risk variant and actually have disease, is described by the expressivity of the variant. As explained above, other factors than only genetic risk factors also must play a role.

Despite the existence of many genetic variants, only a fraction of the genotypic differences contributes to phenotypic variation. The exact locations where in the chromosomes the true causative variants are located is mostly unknown. Testing all of the several millions of common and rare SNPs in a person’s chromosomes would be extremely expensive. But variants that are near each other tend to be inherited together, e.g. people who have a nucleotide A rather than a nucleotide G at a particular location in the chromosome can have identical genetic variants at other SNPs in the chromosomal region surrounding the A. This non-random association between alleles at different loci is termed linkage disequilibrium (LD) and the regions of linked variants are known as haplotypes (www.hapmap.ncbi.nlm.nih.gov) (43). Determining the identity of a common SNP in a haplotype, the tag SNP, uniquely identifies all other linked variants on the same haplotype. Testing an individual’s tag SNPs, enables to identify the haplotypes in the chromosomes. If patients tend to share a particular haplotype, variants contributing to the disease might be somewhere within or near that haplotype. The number of tagSNPs that contain most of the information about the patterns of genetic variation of a genome is estimated to be 300,000 to 600,000, which is far fewer than 11 million common SNPs, and extremely less expensive to
genotype. Thus, the information from the HapMap has been instrumental to map variants associated with various diseases.

The genetic factors associated or actually contributing to the pathogenesis of periodontitis are only limitedly identified. In the last decades research on the genetics of periodontitis has focused on identifying specific SNPs in specific genetic loci as risk factors for AgP and CP. In general, genetic studies using a severe disease phenotype are the most useful to identify the genes involved in the disease. For complex diseases, the strongest phenotypes will be affected most by genetic factors (genetic penetrance is high) and will suffer the least from “noise” from environmental, life style, systemic and other factors (Figures 1-3); in other words, there is more genetic homogeneity in such case populations. Therefore studies employing AgP study populations may have less genetic heterogeneity than CP study populations. The genetic risk loci identified in AgP can then be carried forward for testing as candidate risk loci in CP. Notably, as stressed above, the SNPs identified for periodontitis mostly point to a gene locus of importance, where these SNPs serve only as genetic markers and not particularly as the causative variants.

A large amount of candidate gene association studies have been performed in periodontitis with varying and often contradictory results (13, 26, 46). Small cohort size resulting in lack of power and the lack of replication have been the major problems in the periodontal genetic studies (47). Further, most of the candidate gene studies in periodontitis have not captured the complete genetic information of a particular region of interest. In almost all studies, only one or a few candidate SNP’s instead of complete haplotypes of the genetic locus of interest were genotyped. Furthermore, many studies on genetics in periodontitis are limited because of inadequate phenotype classification of periodontitis and control subjects, as well as not taking into account life style factors such as smoking, possible differences in allele carriership related to gender or the presence of other diseases (47).

5.2 The genome-wide association study (GWAS) approach

A genome-wide association study (GWAS) is a powerful molecular technique to analyze hundreds of thousands or even millions of variations in genomic DNA simultaneously and to determine if any genetic locus is associated with a certain disease phenotype. It is an open ended or hypothesis free approach; no a priori genetic candidate is being investigated. GWAS
analyze SNPs covering the entire human genome with the great majority in regulatory regions and a small part in coding regions. Because of the complex structure of the genome (for example linkage disequilibrium), associated variants are mostly not the causative variants; however the associated variants that are identified are expected to be linked with the causative variants. This can be uncovered in subsequent genetic fine mapping experiments.

Since in a GWAS an enormous amount of SNPs is determined, it is essential to include large numbers of well-characterized (most preferably strong phenotypes, such as AgP rather than CP), homogeneous patient populations (all the same ethnic background) and as good as possible, matched controls. Importantly, the statistical significance level for $p$ is usually set at $<5 \times 10^{-8}$ to correct for extensive multiple testing (48). Typically, on the basis of a GWAS, novel candidate genetic loci or SNPs can be proposed; importantly, the results of a GWAS need to be further validated in an independent cohort with a candidate gene study. Therefore, GWAS are instrumental in discovering novel candidate genes and their possible roles in biological pathways, especially in diseases where the genetic basis is not understood.

Today, more than 1300 GWAS have been performed (www.ebi.ac.uk/gwas/), but only five GWAS have been performed in association with periodontitis; one in AgP and four in CP (49-53). Another GWAS in periodontitis studied SNPs in association with pathologic periodontal pocket depths, rather than in association with a certain form of periodontitis (54). Furthermore, one GWAS examined host genetic risk loci in association with subgingival bacterial colonization (55). Recently, Rhodin et al. (56) reanalyzed the GWAS data of previous studies (50, 55), with a gene-centric and gene set enrichment analyses. We report the validated gene variants associated with periodontitis from these GWAS studies in combination with candidate gene studies below. Notably, the studies differ in methodology and quality, and level of validation.

6. The best replicated and validated genetic factors for periodontitis

We have ordered the results from the literature per chromosome.

**COX2 on Chromosome 1**

Based on robust reasoning on the role of prostaglandins in periodontitis, the gene locus COX2 (alias prostaglandin-endoperoxide synthase-2 [PTGS2]) was proposed as candidate gene (57).
The COX2 SNP rs20417 (haplotype tagged by rs6681231, having near-perfect linkage disequilibrium \(r^2 > 0.95\)) was first identified in a Taiwanese AgP cohort (85 cases versus 153 controls, Table 1) (57) and later also observed in a Dutch-German AgP study population (520 cases versus 1043 controls) (58). So far, these associations have not been replicated in cohorts with same ethnical background.

The association of COX2 SNP rs6681231 was also reported to be associated with CP in Taiwanese (343 cases versus 153 controls) (57). Later, Loo et al. (59) and Li et al. (60) studied the association between CP and the COX2 SNP rs20417 in Chinese patients (280 cases versus 250 controls, 122 cases and 532 controls, respectively) and both studies also reported on an association of this gene with periodontitis, but the association was in an opposite direction as previously suggested. This confirms that SNPs in this genetic locus are related to periodontitis, but that the real causative SNP(s) are not yet identified, only tagging SNPs. No association for the COX2 SNPs has been found in Dutch-German CP patients (58).

COX-2 converts arachidonic acid into prostaglandin H2, which is the precursor of prostaglandin E2 (PGE2). PGE2 mediates pro-inflammatory and anti-inflammatory reactions in many tissues and is also partly responsible for the resorption of the periodontal connective tissues and alveolar bone during the pathogenesis of periodontitis (30). However, whether the identified genetic association is related to altered function of the protein COX-2, and therefore, indirectly could play a role in the pathophysiology of periodontitis is still an open question.

**IL10 on Chromosome 1**

The involvement of the *IL10* genetic locus has already been proposed for a long time (13). Nevertheless, only recently this is sufficiently validated. The associations of *IL-10* SNPs rs61815643 and rs6667202 were identified in a German population of 600 AgP patients and 1441 controls (61). The associations were further tested in validation cohorts of 164 Dutch AgP patients and 1020 controls, and 105 German-Austrian AgP patients and 482 controls of the same countries. The genetic region was confirmed in the Dutch validation cohort, but not in the German-Austrian cohort (61). The IL10 genetic locus as risk factor for CP is currently not yet sufficiently validated.
The *IL-10* genetic region is a very interesting region, it is another example of a pleitropic locus. For example, it has been associated with inflammatory bowel disease and type 1 diabetes (61, 62). The *IL10* gene encodes the anti-inflammatory cytokine interleukine-10 (*IL-10*). *IL-10* can down-regulate the pro-inflammatory immune response of monocytes and macrophages in paracrine and autocrine fashion: *IL-10* is produced by monocytes, macrophages and T cells, and its actions result in reduced expression of the pro-inflammatory cytokines such as *IL-1* and TNF-α.

**IL1 genes on Chromosome 2**

The *IL1* genetic locus encompasses the interleukin (IL)-1 genes *IL1A*, *IL1B* and *IL1RN*. *IL-1A* –889 (rs1800587, in linkage with +4845), *IL-1B* –511 (in linkage with –31), *IL-1B* +3954 (rs1143634, also mentioned in the literature as +3953) and *IL-1RN VNTR* (in linkage with +2018 [rs419598]) have been studied extensively in relation to CP (13). Kornman et al. (63) was the first to report that the combined presence of the minor allele of the *IL-1A* gene at position –889 and the minor allele of the *IL-1B* gene at position +3954 was associated with severity of CP, in particular in non-smoking Caucasians; the latter authors proposed this combination to be the “*IL-1* composite genotype”. Carriage rates of the *IL-1* composite genotype vary across populations, e.g. a low MAF (<5%) was seen in Asian populations as compared to Caucasian populations (13). The *IL-1* composite genotype and the other *IL-1* candidate SNPs are not associated with AgP (61, 64, 65). *IL-1A* and *IL-1B* were originally proposed as candidate genes due to the fact that these gens codes for the pro-inflammatory proteins *IL-1*alpha and *IL-1*beta respectively, which play a major role in upregulation of the immune response and alveolar bone destruction.

Recent systematic reviews on *IL-1* gene polymorphisms in CP-cases and controls in Caucasians, suggested evidence for the minor alleles in *IL1A*, in *IL1B* and the composite genotype to be risk factors for CP (66-68). However, the results from meta-analyses also demonstrated significant heterogeneity among the included studies indicating the possibility that some of the included studies may suffer from a type 1 error, a phenomenon that can occur in small studies. In that respect, a recent letter is useful for the evaluation of systematic reviews on genetic risk factors (69).

**DEFB1 on Chromosome 8**
The gene *DEFB1* was proposed as candidate gene for periodontitis and again firstly investigated in AgP in a cohort of 532 patients and 1472 controls of German-Dutch origin (70). A suggestive association was found between the *DEFB1* SNP rs1047031 and AgP. This association was not replicated in another Dutch-German AgP study population, but positively in a German-Dutch CP cohort (805 cases versus 1415 controls) (70).

The rare allele of *DEFB1* was predicted to impair a microRNA binding site at the 3’-untranslated region (UTR) of this locus. *DEFB1* encodes the antimicrobial peptide beta-defensin 1, which plays a pivotal role in maintaining a healthy status of the mucosal epithelia, including pocket and oral epithelium. However, how the beta-defensin 1 plays a role in the pathophysiology of periodontitis, still remains unclear; and similarly it is unclear whether the genetic polymorphism as just a marker or has a functional effect.

**ANRIL on Chromosome 9**

The associations between AgP and SNPs in the *CDKN2B* antisense RNA 1 (*ANRIL*) locus can be considered as the best replicated genetic associations in periodontitis to date (Table 1) (61, 71-73). The interest for the *ANRIL* locus is based on pleitropy, since this locus was first the best replicated gene locus for cardiovascular diseases, in particular coronary artery disease in Caucasians (71). But this locus is also associated with diabetes and some forms of cancer (73). Thus by a candidate gene approach, *ANRIL* was discovered to be associated with periodontitis.

The first study of the *ANRIL* gene identified a haplotype block tagged by SNP rs1333048 to be associated with generalized AgP (151 cases versus 736 controls) and localized AgP (137 cases versus 368 controls) in a German population (71). This association was replicated in a German-Northern Irish population (130 cases versus 339 controls) (72). In a follow-up study by Schaefer et al (73) the entire *CDKN2BAS* region was genotyped in Dutch (159 AgP cases versus 421 controls) and German (301 AgP cases versus 962 controls) cohorts, and significant replicated associations were found between AgP and SNPs in the *ANRIL* locus. We conclude that the *ANRIL* locus is associated with AgP, however the causative SNP still needs to be determined.
For the ANRIL SNP rs3217992 a trend was found in German CP (P=0.06) (73). However, this association has not been replicated in another CP cohort to date. The latter was concluded, despite that the ANRIL SNP rs10811658 was associated with CP in Dutch as well as in German cohorts, also after adjustment for smoking, diabetes, gender and age, however importantly, after correcting for multiple comparisons the association lost its significance (73).

Interestingly, the ANRIL locus is a regulatory region and does not contain a protein-encoding gene. Further characterization of the molecular function of ANRIL showed, a long-distance regulatory effect on the genetic activity of the CAMTA1/VAMP3 locus, ADIPOR1 and C11ORF10 (74). The CAMTA1/VAMP3 region has also been shown to be associated with increased periodontal pathogen colonization (55, 74).

**GLT6D1 on Chromosome 9**

The gene for a glucosyltransferase (GLT6D1) on chromosome 9 was discovered in the first periodontitis GWAS for AgP (49). The GWAS discovery cohort consisted of 141 AgP patients and 500 controls, and the second GWAS cohort in the same study for replication, consisted of 142 AgP patients and 479 controls, both cohorts with subjects of German origin. From the first GWAS 197 quality-controlled SNPs and from the second GWAS 244 SNPs passed the threshold for association. Only the GLT6D1 SNP rs1537415 remained significant in both AgP GWAS cohorts (Table 1). This association was further validated in a Dutch cohort with localized and generalized AgP (155 cases vs 341 controls) (49).

The GLT6D1 SNP rs1537415 is located within intron 2 of the glycosyltransferase 6 domain containing 1 (GLT6D1) gene. Preliminary functional analyses have shown that the rare allele of SNP rs1537415 results in impaired binding of the transcription factor GATA-3 (49), however its functional role in AgP is not clear.

**Other suggested periodontitis genes based on GWAS**

Divaris et al. performed the first GWAS in CP (50). The study included 761 patients with severe and 1920 with moderate CP, and 1823 periodontally healthy controls of European American origin. No genome-wide significant associations were found (p values were not <5 x 10^{-8}). However, six genetic loci showed suggestive associations with CP (with p values <5 x
10^6). In a validation cohort of 686 individuals (86 controls, 373 moderate and 197 severe periodontitis patients) three of these loci showed concordance with the GWAS results. Moderate CP was associated with SNPs rs7762544 (chromosome 6, closest gene NCR2) and rs3826782 (chromosome 19, closest gene EMR1), and severe CP with SNP rs2521634 (chromosome 7, closest gene NPY) (50).

Teumer et al. also used a GWAS to study genetic loci (17 million genetic variations) in association with CP in two German populations (51). The first cohort consisted of 670 severe, 1188 moderate CP patients and 1507 controls, and the second cohort consisted of 111 severe, 247 moderate CP patients and 309 controls. No genome-wide significant association was found for CP (p values were not <5 x 10^-8). However, interestingly, they reported that 25% of CP is explained by genetic factors, and even 34% of variance for mean approximal periodontal attachment loss in individuals younger than 60 years old, was explained by genetic factors. This is in agreement with the previous reports estimating heritability for mean attachment loss in patients with CP (21, 24).

Shaffer et al (54) performed a GWAS in non-Hispanic Caucasian adults in association with periodontal pocket depth. No significant genome-wide association was found. The latter study made no distinction for their patients having either AgP or CP, which made the study results also difficult to interpret and compare to others. The AgP phenotype is the more severe phenotype of periodontitis, and has a stronger genetic component in its multi-causality.

The most recent GWAS including a validation cohort was performed on 2760 Japanese CP cases and 15158 controls (53). No SNPs passed the high level of significance set for GWAS (p<5 x 10^-8), but two genetic loci were proposed for further replication and validation studies: KCNQ5 on chromosome 6 and GPRI41-NME8 on chromosome 7.

In another approach, subgingival infection patterns with oral bacterial species in patients with CP and controls have been investigated as outcome parameter in a GWAS (55). A total of 1020 Caucasian Americans were included: 416 subjects diagnosed with healthy/mild periodontitis, 415 with moderate and 189 with severe periodontitis. Subgingival plaque samples were analyzed by DNA-DNA hybridization. Subjects with “high” bacterial colonization of “red” complex species: *P. gingivalis, T. forsythia,*
Treponema denticola; “orange” complex species: Prevotella intermedia, Prevotella nigrescens, Fusobacterium nucleatum, Campylobacter rectus, and Aggregatibacter actinomycetemcomitans were compared with the non-“high” colonizers. No genome-wide association with “high” bacterial colonization or CP was found. However, suggestive evidence was found for 13 loci including KCNK1, FBXO38, UHRF2, IL33, RUNX2, TRPS1, CAMTA1 and VAMP3 (55); these results could not be validated in an African American replication cohort (N=123). However, a recent new study with a candidate gene approach, testing possible SNPs in genes associated with coronary artery disease (CAD) in Northern European AgP patients and controls, validated the involvement of CAMTA1 and VAMP3 (75).

The recent gene-centric and gene set enrichment analyses on existing GWAS data have offered a promising approach to identify novel periodontitis genes with less demanding cutoffs for multiple testing (54). Four genes (NIN, ABHD12B, WHAMM and AP3B2) were associated with severe CP, and two genes (KCNK1 and DAB2IP) with periodontal pathogen colonization. Furthermore, some other genes were proposed for moderate CP. Several of these associations confirmed the suggestive associations of the earlier studies, and need to be replicated in independent cohorts.

Although to date GWAS have not identified any genome-wide associations with CP, several suggestive genetic loci have been identified (49, 51, 55). These results have created novel candidate genes and advanced our understanding of genetic factors in the pathophysiology of CP. Nevertheless still open questions remain. GWAS have not been able to validate most of candidate gene associations (exception IL10), contrasting with previous meta-analyses for example on IL-1 polymorphisms (60, 67).

Other suggested periodontitis genes based on the candidate gene approach

Many complex diseases, such as rheumatoid arthritis (RA), Crohn’s disease (CD) and ulcerative colitis (UC) (both CD and UC are termed inflammatory bowel disease (IBD)), type 2 diabetes, cardiovascular diseases and other inflammatory diseases, share genetic risk factors (41, 62). Pleiotropy has been found for 13.2-18.6% of disease associated genes and in 4.6-7.8% of SNPs associated with disease or disease traits (www.genome.gov). The sharing of genetic variants for multiple diseases and phenotypes may generate new candidate genes for
other diseases and can also be incorporated into the future research on genetic risk variants in aggressive and CP (42).

With the approach as described above, new candidate genes for AgP were explored by considering known RA and systemic lupus erythematosus (SLE) genes (76). In the latter study, variants within the IRF5 and PRDM1 genes were suggested to be associated with AgP, but not robustly validated; the proteins encoded by these genes play a role in interferon-beta signaling (76). In a similar fashion, recently new candidate genes for AgP were identified by exploring the well-established risk genes for coronary artery disease (CAD) (75). The CAD-associated gene PLG (plasminogen), was significantly associated with AgP in both an explorative cohort and a validation cohort.

Nevertheless, many genetic variants are not shared between inflammatory diseases or possibly not even between AgP and CP. For example, the genome-wide association of GLT6D1 with AgP has not been found in any other genetic disease or trait not even in CP (42).

7. Discussion and conclusions

Periodontitis has a multi-causal etiology, in which genetic factors play a role. The various proposed causes for periodontitis always play a role simultaneously, but the relative contribution of each of these, varies from case to case. In young individuals often with AgP, a stronger contribution from genetic factors is apparent, while in older individuals often with CP, the relative contribution of the established subgingival biofilms (environmental factors) and life style factors (e.g. smoking, stress, diet), play a more dominant role in the phenotype of the disease. Nevertheless always some genetic susceptibility is present, for CP this is estimated at 25%. Periodontitis is therefore a complex disease, i.e. it behaves in a nonlinear fashion. Actually, the disease progression rate fluctuates, where the disease sometimes moves into an aberrant state of host response and then swings back into a resolving state; between these zones, an accommodation (settlement) zone is present where essentially no differences are found for immunological parameters between cases with periodontitis and healthy controls. The genotype determines part of this fluctuation described and the extent of it.

Like many chronic complex diseases, a multitude of genes (>100) is most likely involved; the disease is therefore polygenic. Currently, only a fraction of susceptibility genes is robustly
identified, namely \textit{GLT6D1}, \textit{ANRIL} and \textit{COX2}, \textit{IL1} and \textit{IL10} have repeatedly been associated with periodontitis (49, 57, 58, 61, 67, 71-73). Most of the associations have been reported for AgP emphasizing the importance of genetic factors in the pathobiology of this severe periodontitis phenotype. Interestingly, polymorphisms within \textit{ANRIL} and \textit{COX2} tended to be associated also with CP (57, 59, 60, 73). Further, SNPs within the regulatory regions of the \textit{IL-10} gene have shown suggestive associations with AgP, and SNPs in \textit{DEFB1} with both AgP and CP (61, 70). Probably, the identified SNPs are “genetic markers” and are not the true causative variants for both AgP and CP.

The identified genetic risk variants for periodontitis are all located within introns (\textit{ANRIL} and \textit{GLT6D1}) or regulatory regions (\textit{COX-2, IL-10} and \textit{DEFB1}) (42). This is in line with other complex diseases; the majority of common human genomic variants that are genome-wide associated with over 400 complex diseases and traits are located within regulatory and intronic regions and not within coding regions (77). Genetic variations of intronic and regulatory regions may lead to subtle changes in the expression of associated coding regions and may affect the quantity of the transcription and subsequent protein product.

All individuals harbor millions of genetic variants (SNPs, insertions, deletions, repeats, etc.). The majority of these variants most likely have no effect, while the minority may have some function, mainly physiological and not pathological, and are mainly responsible for the “normal” phenotypic differences between individuals. However, some combinations of the inherited variants can make individuals susceptible for certain traits or diseases, always in combination with unfavorable environmental, lifestyle and other factors; in particular if the latter are unfavorable for the host resistance, the genetic penetrance of inherited variants increases (77) and the host defense may move into an aberrant immune state (Figure 4). Another important aspect, essentially not addressed in this chapter, is the potential role of epigenetics in the pathobiology of periodontitis (14). This new and emerging field will yield in the years to come new valuable information in relation to the susceptibility of periodontitis.

For the disease periodontitis, we continue the search for genetic risk factors. This may help to better understand the disease, thus to be able to identify pathobiological pathways where genetic factors, environment, lifestyle and other factors are interconnected. To date, there are several modestly large cohorts of patients and controls available (Northern Europe, USA,
Japan), future studies need to concentrate on larger sample sizes in multiple ethnic populations for discovery and validation, and to use (mathematical) modelling (for example references (37, 78)) of the various causative factors to further explore the complexity of periodontitis.

Acknowledgement
Authors B.G.L. and M.L.L. are supported in part by a grant from the University of Amsterdam, the Netherlands, for the focal point ‘Oral infection and inflammation’.

Disclosures
No potential conflicts of interests relevant to this review are reported.
References


